

CFTRI-MYSORE



9666

Biophysical chem.





TLMS



BIOPHYSICAL CHEMISTRY

Volume I

BIOPHYSICAL CHEMISTRY

JOHN T. EDSALL

*Biological Laboratories,
Harvard University
Cambridge, Massachusetts*

JEFFRIES WYMAN

*Middle East Science
Cooperation Office,
UNESCO, Cairo, Egypt*

VOLUME I

THERMODYNAMICS, ELECTROSTATICS, AND THE
BIOLOGICAL SIGNIFICANCE OF THE
PROPERTIES OF MATTER



1958

ACADEMIC PRESS INC • PUBLISHERS • NEW YORK

9666

COPYRIGHT 1958 BY ACADEMIC PRESS INC.

ALL RIGHTS RESERVED

NO PART OF THIS BOOK MAY BE REPRODUCED IN ANY FORM
BY PHOTOSTAT, MICROFILM, OR ANY OTHER MEANS,
WITHOUT WRITTEN PERMISSION FROM THE PUBLISHERS.

ACADEMIC PRESS INC.

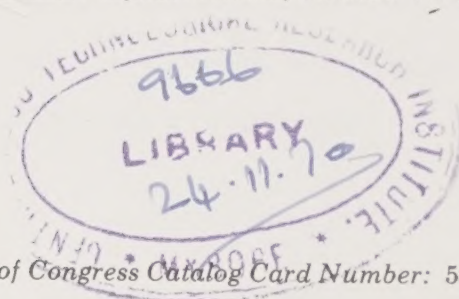
111 FIFTH AVENUE
NEW YORK 3, N. Y.

United Kingdom Edition

Published by

ACADEMIC PRESS INC. (LONDON) LTD.

BERKELEY SQUARE HOUSE, LONDON W. 1



Library of Congress Catalog Card Number: 58-5714

First Printing, 1958
Second Printing, 1962
Third Printing, 1966

GUC:(E)
J8

PRINTED IN THE UNITED STATES OF AMERICA

CFTRI-MYSORE



9666

Biophysical chem.

Preface

The term Biophysical Chemistry is relatively new, but much of the subject matter has been studied intensively for more than a generation. In the present volume we deal first with the biological significance of the properties of matter, laying special emphasis on the chemical elements, on such simple compounds as water and carbon dioxide, and on certain very complex molecules, notably the proteins. These subjects are covered in the first three chapters, except for the discussion of carbon dioxide, which is placed in Chapter 10, following the chapters on acid-base equilibria. Thermodynamics and electrostatics, and their significance for biochemistry, are treated at length in Chapters 4, 5, and 6. We have aimed to start from elementary principles, and to develop those subjects coherently until problems on the frontiers of modern knowledge have been reached. In Chapter 7, electrical conductance is treated much more briefly, but we hope adequately for our purposes. Acid-base equilibria, including the foundations of the concept of pH and the special characteristics of polyvalent acids including proteins, are treated in great detail in Chapters 8 and 9. The discussion in these chapters lays the foundation for the extensive general treatment of molecular interactions in the last chapter of Volume I.

Volume II will deal with the physical chemistry of biological macromolecules and their study by such methods as osmotic pressure, light scattering, diffusion, sedimentation, viscosity, dielectric dispersion, double refraction of flow, and electrophoresis. It will conclude with several chapters on the physical chemistry of blood.

We have assumed that the reader is familiar with the fundamentals of differential and integral calculus and has some knowledge of organic and physical chemistry and general biology. Advanced knowledge of these subjects is not essential, but will certainly enable the reader to make more rapid progress in the more difficult chapters. We hope that the book will be found valuable by graduate students with the requisite background, and by some advanced undergraduates also, and that it may be found useful as well by active investigators in chemistry, biology, and biochemistry.

We have deliberately aimed to deal with a limited number of topics with some depth, rather than to include all subjects which might be included in a treatise on Biophysical Chemistry. We have striven in many places to build a bridge over which the reader may pass from the elemen-

tary concepts of a subject to advanced and difficult problems. Thus the discussion of electrostatics in Chapter 5, beginning with Coulomb's law and Gauss's law, proceeds directly to a derivation of Poisson's equation for systems with spherical symmetry, and thence to the development of the Debye-Hückel theory of interionic forces. Some passages in this chapter, such as that on Image Charges, may be passed over on the first reading if desired, and taken up later. In the discussion of ion-dipole interactions, a relatively simple model is presented which gives the correct form of the equations for the influence of ionic strength and dielectric constant upon the activity coefficients of dipolar ions; this serves as a prelude to the discussion of the more elaborate models treated by Kirkwood. Likewise, in the discussion of the dielectric constants of polar liquids in Chapter 6, a simple model is used which reproduces many of the major features of the theories of Onsager and Kirkwood but with a much less complicated mathematical development. The more rigorous theories are then presented. The treatment of microscopic and macroscopic dissociation constants in Chapter 9 is developed for the general case, and concretely illustrated by detailed calculations for glutamic acid and cysteine.

No apology is needed for the strong emphasis on proteins in this book. Their supreme importance in biological systems, and the inherent fascination of the problems encountered in unraveling their structure, would fully justify the length of the discussion we have given. On the other hand, we regret having omitted a discussion of nucleic acids, which it had been our intention to include. This omission would have been remedied if we had had unlimited time, but we felt that it was better to discuss first the things we knew most intimately. We had also hoped, originally, to include a discussion, in either Volume I or Volume II, of reaction kinetics and the mechanism of enzyme action. This still remains our hope for the future, if we should ever be able to prepare a supplementary third volume.

We both wish to acknowledge our debt to our teacher, the late Lawrence J. Henderson, whose thinking has deeply influenced many biologists. The effect of the ideas developed in his philosophically provocative book "The Fitness of the Environment" will be apparent in Chapters 1, 2, and 10 of the present volume, and that of his Silliman lectures, "Blood: A Study in General Physiology," will be found in our discussion of the physical chemistry of blood in Volume II. In the study of proteins in general and of blood plasma proteins in particular, we are deeply indebted to the late Edwin J. Cohn. The monograph "Proteins, Amino Acids and Peptides," by Cohn, Edsall, and others, which appeared in 1943, bears an obviously close relation to some parts of the

present book. However, what we have written is by no means to be considered as a revised version of that monograph, which not only discussed general principles but provided comprehensive tabulations of experimental data. In the present book our emphasis is on the broad general principles and fundamental concepts. We have made extensive use of experimental data to illustrate the principles, but no attempt has been made to compile data in comprehensive fashion. We have considered it vitally important to give illustrations of the general principles from actual experimental data—as for instance in the calculations of the free energy changes in biochemical processes which are discussed in the section on “Mass Law” in Chapter 4—and the particular examples chosen do, we believe, represent experimental work of outstanding quality. Many other experimental studies of equal quality, however, could readily have been chosen.

Likewise the lists of references cited at the ends of the chapters are not intended to be comprehensive, but rather to guide the reader into further interesting and helpful reading on the subjects we have discussed. Many of the general references cited are not specifically mentioned in the text, so that the reader should examine the references carefully. Sometimes—for instance in Chapters 3 and 9—when dealing with subjects in which current developments are occurring with great rapidity, we have provided extensive references to the original literature. In some other chapters only a few references are given, and these are mostly of a broad general character, with only a few original papers of particular interest cited in addition.

This book has grown out of a course at Harvard in the physical chemistry of biochemical systems, originally given by one of us, starting more than twenty-five years ago. Later, for a number of years, we gave the course jointly. Finding no book that followed our presentation at all closely, we were led to the preparation of extensive mimeographed notes for the students, and eventually to the writing of this book. The course, now known as Biology 184, is still being given by the other of us; but the book has grown in scope, far beyond the level of what can be covered in a half-year course, or contained within a single volume.

We are indebted to many friends and colleagues for help with this book. G. Evelyn Hutchinson and Kirtley F. Mather have read the manuscript of Chapters 1 and 2 and offered helpful suggestions; Barbara W. Low has read Chapter 3; Alexander Rich has been particularly helpful in advising us concerning the discussion of collagen in the same chapter. J. G. Kirkwood and C. P. Smyth have also examined much of Chapters 5 and 6, and other colleagues have examined other portions of the book. We are indebted to J. L. Oncley for many discussions of problems dealt

with in this book, and especially for a number of valuable suggestions regarding the chapter on Dielectric Dispersion which will appear in Volume II. Obviously these colleagues bear no responsibility for the opinions we have expressed, or for the errors which undoubtedly remain.

We are especially grateful to Mr. Fred Rothstein, who has read every chapter in detail, often in several successive drafts, and has pointed out numerous errors and obscurities which might otherwise have been overlooked. His point of view, which is that of an advanced graduate student, has been particularly helpful to us in our endeavor to express difficult concepts with maximum clarity, but without evasion of the difficulties. We are not so sanguine as to believe that we have fully achieved these objectives, or that all errors and obscurities have been eliminated, but Mr. Rothstein's criticisms have certainly helped us immeasurably.

For the extensive and arduous work involved in typing many revisions of the manuscript we are indebted to Barbara T. Coffin, Katharine Reynolds, and especially to Laura L. Williams. The preparation of the index, which was a formidable undertaking, has been carried through by Margaret S. Little.

One of us (J.T.E.) is deeply indebted to the John Simon Guggenheim Memorial Foundation for a special fellowship during the years 1954-56 which provided not only support for his research but also extra facilities which did much to make his share in the writing of this book possible. Helpful support was also received from the National Science Foundation in the form of a research grant which also aided in the opportunity for ripening many of the ideas presented here.

In the production of the book we have received constant and understanding help from the staff of the Academic Press.

JOHN T. EDSALL
JEFFRIES WYMAN

November, 1957

In the second printing of this book we have taken the opportunity to correct a few errors and to revise extensively the discussion of ribonuclease, and the brief comments on hemoglobin, on pages 97-99. A supplementary list of references relating to this material will be found on p. 662.

April, 1962

Contents

Preface.....	v
Symbols.....	xiv
1. Biochemistry and Geochemistry.....	1
Distribution of the Chemical Elements in Living Organisms.....	3
The Relative Abundance of the Chemical Elements in the Universe.....	7
The Structure of the Earth and Its Crust.....	8
Early History of the Earth.....	11
The Ocean.....	17
The Central Role of Carbon in Biology.....	23
References.....	25
2. Water and Its Biological Significance.....	27
Heat Capacity of Water and Other Substances.....	37
Heat of Vaporization.....	38
Latent Heat of Fusion.....	39
The Expansion of Water in Freezing.....	40
Surface Tension.....	42
Dielectric Constant of Water.....	43
Solubility.....	45
References.....	45
3. Problems of Protein Structure.....	47
Proteins; Some General Considerations.....	48
Amino Acids and Peptides as Dipolar Ions.....	54
Characteristics of the Amino Acids Derived from Proteins.....	56
Amino Acid Analysis of Proteins.....	65
Polypeptide Chains and Their Presence in Proteins.....	74
Determination of Arrangement of Peptide Chains and Their Linkage within Protein Molecules.....	80
The Number of Peptide Chains in the Molecule, and the Nature of the Terminal Groups.....	82
Cross-Linkages between Peptide Chains: Disulfide and Phosphate Cross-Linkages.....	85
Disulfide Linkages.....	86
Phosphate Cross-Linkages.....	88
Other Cross-Linkages.....	89
Sequence of Amino Acid Residues in Peptide Chains.....	96
Disulfide Linkages in Insulin and Ribonuclease.....	99
Spatial Configurations of Polypeptide Chains.....	104
Silk Fibroin.....	106
β -Keratin.....	110
α -Keratin and Synthetic Polypeptides: the α -Helix and Other Possible Helical Structures.....	115
Collagen.....	115

Amino Acid Composition.....	116
Heat Shrinkage.....	117
The Structural Pattern—Evidence from X-Ray and Electron Microscope Studies.....	118
Collagen Structure.....	118
The Stability of Helical Structures.....	123
The Significance of Disulfide Bonds; Cross-Linkages and Loops in Peptide Chains.....	125
Possible Stereochemistry of Insulin.....	126
Stereochemistry and Enzyme Activity of Ribonuclease.....	128
Helical and Nonhelical Regions in Globular Proteins.....	129
References.....	132
4. Thermodynamics.....	136
Introduction.....	136
Systems and Phases.....	136
Temperature.....	138
Dimensions of Temperature.....	140
First Law.....	142
Second Law.....	146
Irreversible Processes.....	157
Maximum Work, Equilibrium, and Free Energy.....	159
Heat Content, or Enthalpy.....	162
Expressions for the Change of the Fundamental Thermodynamic Functions..	162
Partial Molal Quantities.....	170
Chemical Potentials.....	174
Phase Rule.....	176
Activities.....	181
Ideal or Perfect Solutions and Raoult's Law.....	184
Activity Coefficients and the Choice of Standard States.....	189
Activity and Chemical Potential of a Nonvolatile Solute from the Vapor Pres- sure of the Solvent.....	195
Equilibrium between Phases; Activities in Relation to Distribution Coefficients and Solubilities.....	198
Activity of Strong Electrolytes.....	200
Mass Law.....	201
The Malate-Fumarate Equilibrium.....	204
Peptide Bond Synthesis in the Reaction between Benzoyl-L-Tyrosine and Glycinamide.....	207
The Standard Free Energy of Hydrolysis of Adenosine Triphosphate.....	210
Systems Involving Other Variables.....	217
Elastic Bodies.....	217
Systems in Gravitational or Centrifugal Fields.....	219
Statistical Interpretation of Entropy.....	224
Entropy of a Perfect Crystal at 0°K.....	228
Entropy Change on Mixing of Components in an Ideal Solution.....	229
The Residual Entropy of Ice.....	231
Standard Free Energies of Formation and Their Use in Determining Thermo- dynamic Equilibria.....	232
The Fumarate-Malate Equilibrium.....	235

The First Ionization of Carbonic Acid.....	235
Standard Free Energy of Peptide Bond Synthesis.....	238
References.....	238
5. Electrostatics: Its Application to Polar Molecules and Ionic Solutions.....	241
Definition of the Potential.....	243
Gauss's Law.....	246
Poisson's Equation for Regions Containing a Space Charge.....	249
Free Energy of a Charged Sphere.....	252
The Potential and Energy of a Dipole.....	254
Image Charges.....	258
The Salting-Out Effect.....	263
Salting Out of Proteins.....	275
Effects of Dipolar Ions.....	278
Debye's Theory of Salting Out by Redistribution of Solvent Molecules around the Ions.....	278
Ionic Interactions and the Debye-Hückel Theory.....	282
Ion-Dipole Interactions.....	296
Experimental Studies on the Solubility of Dipolar Ions, and Their Interaction with Ions.....	306
Solubilities and Activity Coefficients in Water and Other Media.....	308
Solubilities in the Presence of Salt.....	310
References.....	320
6. Dielectric Constants and Their Significance.....	323
Introduction to Dielectric Constants and Dipole Moments.....	323
Dielectric Polarization in Relation to Molecular Properties.....	328
The Debye Theory of the Dielectric Constant.....	328
Application of the Debye Theory.....	335
The Breakdown of the Debye Theory in Polar Liquids.....	341
Modifications of the Debye Theory.....	342
The Onsager Theory.....	347
Kirkwood's Theory.....	356
Polar Liquids and Dipolar Ions.....	364
Dielectric Constants and Dipole Moments of Dipolar Ions in Solution.....	367
Methods of Measuring Dielectric Constant.....	378
References.....	385
7. Conductivity of Electrolytes.....	386
Conductance and Resistance.....	386
Electrolytes as Conductors.....	387
Faraday's Law.....	388
Equivalent Conductivity.....	389
Ion Mobilities.....	390
Ion Conductances.....	393
Transference Numbers.....	393
Hittorf Method of Determining Transference Numbers.....	394
The Moving Boundary Method of Determining Transference Numbers.....	395
Theoretical Considerations Regarding Equivalent Conductance; Effects of Interionic Forces.....	401
Measurement of Conductivity.....	404
References.....	404

8. Acid-Base Equilibria	406
The Nature of Acids and Bases.....	406
The Concept of pH ; a Preliminary Statement.....	413
Calculations of pH in Systems of Acids of Known Acid Strengths....	414
The Titration of a Weak Acid with a Strong Base; Buffer Action.....	417
The Experimental Determination of Acidity Constants (K_A Values).....	420
The Conductivity Method.....	421
The Indicator Method for Determining Relative Strengths of Acids....	423
Electromotive Force Measurements and pH Determination.....	429
Standard Potentials of Half-Cells, Choice of Conventions.....	434
Evaluation of pK_A from Cells without Liquid Junction.....	436
Cells Containing Liquid Junctions.....	437
The Liquid Junction Potential.....	439
Variation of pH with Ionic Strength for Acid-Base Pairs of Different Charge Types.....	441
Choice of a Standard for Calibrating pH Measurements.....	444
The Glass Electrode.....	449
Determination of pK_A Values and Related Thermodynamic Functions in Relation to Structure.....	450
Effect of Neighboring Charged Groups and Dipoles on pK_A Values.....	457
Other Acidic Groups of Biochemical Interest.....	463
Phenolic Hydroxyl Groups.....	463
Imidazole Groups.....	468
Sulfhydryl Groups.....	468
The Guanidinium Group.....	469
The Peptide Linkage.....	470
Purines, Pyrimidines, Nucleosides, and Nucleotides.....	470
Effects of Variation in Dielectric Constant of Solvent on Relative Strength of Acids of Different Charge Types.....	471
Problems.....	473
References.....	475
9. Polybasic Acids, Bases, and Ampholytes, Including Proteins	477
Dibasic Acids.....	477
Polyvalent Acids: General Relations.....	487
A Specific Case: Glutamic Acid.....	494
Spectroscopic Determination of Microscopic Constants: The Ionization of Cysteine.....	496
Isoelectric Points of Amphoteric Substances.....	504
Isoelectric Points of Polyvalent Ampholytes.....	507
Equilibrium between Different Ionic Forms in Polyvalent Acids and Ampholytes: The Equation of Linderstrøm-Lang.....	510
Electrostatic Effects on Ionization in Polybasic Acids. The Charged Sphere Model.....	512
Formulation in Terms of Association Constants.....	518
Acid-Base Equilibria in β -Lactoglobulin Solutions.....	522
Ovalbumin.....	530
Serum Albumins.....	532
Ribonuclease.....	536
Hemoglobin.....	537
Other Proteins.....	540

Synthetic Polyelectrolytes, Including Polypeptides.....	540
Effects of Internal Hydrogen Bonding on Titration Curves.....	545
Calculations with a Dielectric Sphere Model Containing Fixed Charges.....	546
References.....	547
10. Carbon Dioxide and Carbonic Acid.....	550
Carbon Dioxide in Nature.....	550
Structural Considerations.....	553
The Fundamental Equilibria in Systems Containing Carbon Dioxide and Carbonic Acid.....	555
Absorption Coefficients of Carbon Dioxide in Water.....	555
The Reversible Hydration of Carbon Dioxide and the Ionization of Carbonic Acid.....	557
Carbon Dioxide Dissociation Curves.....	561
The Formation of Carbamates.....	571
Rates of the Reactions $\text{CO}_2 + \text{H}_2\text{O} \rightleftharpoons \text{H}_2\text{CO}_3$	578
The Velocity of the Dehydration Reaction: $\text{H}_2\text{CO}_3 \rightarrow \text{H}_2\text{O} + \text{CO}_2$	580
The Hydration Reaction: $\text{CO}_2 + \text{H}_2\text{O} \rightarrow \text{H}_2\text{CO}_3$	583
Carbonic Anhydrase.....	587
References.....	589
11. Some General Aspects of Molecular Interactions.....	591
Some Methods of Measuring Binding.....	594
Equilibrium Dialysis.....	594
Ultrafiltration.....	595
Spectroscopic Measurements.....	596
Electromotive Force Measurements.....	596
Ultracentrifuge Measurements.....	598
Electrophoresis.....	599
Effects of pH Changes When Salts Are Added: Isoionic Points.....	599
Isoelectric and Isoionic Solutions.....	603
Some Fundamental Considerations Regarding Certain Types of Binding.....	605
Interactions of Bound Groups.....	606
Competition Effects.....	607
Interactions between Different Kinds of Bound Groups.....	607
Formation of Chelate Complexes.....	608
Binding by a Set of Equivalent and Independent Groups.....	610
The Logarithmic Plot: Bjerrum's "Formation Function".....	615
The Reciprocal Plot.....	616
The Scatchard Plot of $\bar{v}/(A)$ against \bar{v}	617
Relation of Equations for Binding to Those of Enzyme Kinetics.....	620
General Equation for Binding by a Molecule or Ion with n Sites Available for Combination.....	623
Binding by an Equivalent Set of Groups with Interactions between Them.....	626
Complexes of Biochemical Substances with Metallic Ions: Some General Considerations.....	638
Electrostatic Effects on Binding of Ions.....	645
Effects of Competition between Different Ligands for the Same Binding Site.....	651
Linked Functions.....	653
References.....	660

Symbols

a	the "collision diameter" for inter-ionic interactions (Chapter 5)	I	electric current (Chapters 6 and 7)
a_i	activity of component i	K	with appropriate subscript, an equilibrium constant
b	the radius of a spherical ion (especially in Chapters 5, 9, and 11)	K_s	a salting-out constant (Chapter 5, p. 272 ff.)
C_i	molar concentration of component i	k	Boltzmann's constant = R/N = 1.380×10^{-16} erg per degree
C_p	heat capacity at constant pressure	k	with subscripts, an intrinsic ionization constant of an individual group (Chapter 9)
C_v	heat capacity at constant volume	k , or k_{assoc}	an association constant (Chapter 11)
D	dielectric constant	l	length (Chapter 4, Eq. 127 <i>et seq.</i>)
D_0	dielectric constant of a substance taken as the standard state	L_1, L_2 , etc.	products of dissociation constants, defined on p. 492
D	electric displacement (Chapter 6, Eq. 3)	L_1^*, L_2^* , etc.	products of association constants, defined on p. 624; see also the constants L_{ij}^* , defined on p. 657
E	internal energy (especially in Chapter 4); numerical value of electric field strength (Chapters 5 and 6); electromotive force (Chapter 8)	m	molality (moles of solute per kg. solvent)
E	electric field strength considered as a vector	N	Avogadro's number, 6.024×10^{23}
emf	electromotive force	N_i	mole fraction of component i
esu	electrostatic units	n	number of binding sites for ligands, in a macromolecule; or coordination number of an ion (Chapter 11)
F	Gibbs free energy	n_i	number of molecules of component i per unit volume (Chapters 5 and 6)
F_e	electrical free energy	n_i	refractive index of component i (Chapter 6)
F	internal field intensity (Chapter 6)	P	total pressure in a system
F	the Faraday equivalent of electric charge = 96,500 coulombs	P	a factor employed in certain equations of the Debye-Hückel theory; defined on pp. 290-291
f	tension (Chapter 4, Eq. 127 <i>et seq.</i>)	p_i	vapor pressure of component i
f_i	activity coefficient of component i (see Chapter 4); commonly taken as ratio of activity to mole fraction, but sometimes in other concentration units; see also γ_i	$pH \equiv -\log a_H$	See discussion of this concept in Chapter 8, pp. 437-449
G_1, G_2 , etc.	titration constants for acid dissociation (Chapter 9, p. 481 ff. and p. 492 ff.)	pH_I	pH of isoelectric point (pp. 505, 509)
g_1, g_2 , etc.	titration constants for association reactions (Chapter 11, p. 625)	Q	heat absorbed (Chapter 4)
h	mean number of protons removed from a polyvalent acid (Chapter 9)		

Q	a factor defined in Chapter 5, p. 291	γ_i	activity coefficient of component i , usually the ratio of activity to molar concentration or molality; see also f_i
Q	Henry's law coefficient for carbon dioxide (Chapter 10)	ϵ	the proton charge = 4.802×10^{-10} electrostatic units
Q	a function defined in Chapter 11, Eq. (38), p. 631	ϵ	molar extinction coefficient (Chapter 8, pp. 424-428)
Q'	apparent heat of ionization in kcal/mole (Chapter 9, Eq. 69)	κ	a reciprocal length, important in the Debye-Hückel theory, proportional to the square root of the ionic strength; defined on p. 287
q'	Henry's law coefficient for carbon dioxide (Chapter 10, p. 563)	κ_1, κ_2 , etc.	intrinsic association constants, defined in Chapter 11, p. 624
R	the molar gas constant = 1.987 calories per degree per mole = 4.183×10^7 ergs per degree per mole	Λ	equivalent conductance (Chapter 7, p. 390)
R	a distance; sometimes the distance between the two poles of a dipole (especially in Chapter 5)	Λ_0	equivalent conductance at infinite dilution
R	electrical resistance (Chapter 7)	λ	the slope of an association curve, defined in Chapter 11, Eq. (41), p. 635
r	distance in general; sometimes the distance between two charged bodies, or between a charge and a dipole	μ	dipole moment
S	entropy	\mathbf{u}	dipole moment, considered as a vector (Chapter 6)
T	absolute temperature in $^{\circ}\text{K}$	μ_i	chemical potential of component i (Chapter 4, p. 174 and elsewhere)
t_+ , or t_-	transference number of a cation or anion (Chapter 7, p. 393)	ν	number of ions present in a salt molecule (Chapter 5)
u_+ , or u_-	electrical mobility of a cation or anion (Chapter 7, p. 390 ff.)	$\bar{\nu}$	mean number of ligand molecules or ions bound by a macromolecule, or by some other molecule or ion under study (Chapters 9 and 11)
V	volume	ϕ	the azimuthal angle in polar coordinates (Chapter 5, Fig. 1)
\bar{V}_i	partial molal volume of component i	ψ	electrical potential
\bar{v}_i	partial specific volume of component i	ρ	density; sometimes density of electrical space charge (see p. 249 ff.)
W	work (in Chapter 4 and elsewhere)	θ	the polar angle in polar coordinates (Chapter 5, Fig. 1)
w	an electrostatic factor defined in Chapter 9, Eq. 53, p. 514	θ	empirical temperature (Chapter 4)
Z	valence (or net charge in proton units) of component i^*	Ω	number of configurations corresponding to a given macroscopic state (Chapter 4, p. 226 ff.)
\bar{Z}	mean net charge of component i	ω	ionic strength (Chapter 5, Eq. 52, p. 273)
α	fraction of the total quantity of an acid-base pair which is in the form of the conjugate base (Chapter 8, Eq. 19, p. 415)		
κ	electrical conductance (Chapter 7, Eqs. 1 and 3)		
κ	specific conductance (Chapter 7, Eq. 3)		

* In Chapter 7, valence of an ion is denoted by z_i .

Chapter 1

Biochemistry and Geochemistry

Carbon is unique among the elements in the number and variety of the compounds which it can form. Over a quarter of a million have already been isolated and described, but this gives a very imperfect idea of its powers, since it is the basis of all forms of living matter. Moreover, it is the only element which could occupy such a position. We know enough now to be sure that the idea of a world in which silicon should take the place of carbon as the basis of life is impossible; the silicon compounds have not the stability of those of carbon, and in particular, it is not possible to form stable compounds with long chains of silicon atoms.—N. V. Sidgwick, "The Chemical Elements and Their Compounds," Vol. I, p. 490.

In this book, we shall deal with the physical chemistry of biochemical systems derived from living organisms, or functioning within such organisms.

For the most part, we recognize what we mean when we speak of a living organism. Living things are always chemically complex; they contain certain very large organic molecules, such as the proteins and nucleic acids, the synthesis of which no chemist has yet achieved in the laboratory. They contain these substances, not in a random mixture, but in a highly organized system. They consume energy, not only in order to move or to grow, but in order to maintain themselves in a steady state—that is, they metabolize. By one fashion or another, they reproduce their kind. These characteristics—complex organized structure, metabolism, reproduction—may be taken as a convenient set of criteria for the general class of living organisms. Inevitably there are borderline systems, such as certain viruses, about which it becomes essentially meaningless to argue whether they are living organisms or not. Viruses are complex structures, and they reproduce, but their metabolism—at least over long stages of their existence—is virtually nil. In practice, however, the existence of these transition states between the living and the nonliving need not disturb us. We shall be concerned with living organisms, and the simpler systems derived from them, such as enzymes and their substrates, as systems which can be studied and analyzed by the methods of physical chemistry.

The principles involved are for the most part the same as those with which the physical chemist dealing with simpler systems is familiar, but

the special properties of the complex structures produced in the living organism will lead us to emphasize some particular problems. The high degree of organization of biochemical systems means that a modification of one of the variables that determines the behavior of the system sets up interactions that produce changes in other variables. A simple and famous example is found in the study of solutions of the protein hemoglobin—the iron-containing, oxygen-transporting protein of the red blood cells. In most chemical systems containing liquids at equilibrium, one may equilibrate a liquid phase with a gas phase containing oxygen and carbon dioxide, and vary the partial pressure of either gas without appreciably affecting the amount of the other which the liquid can take up. If a solution of hemoglobin is in equilibrium with a gas phase containing both these gases, however, it is found that an increase in the partial pressure of carbon dioxide causes a marked release of oxygen from the solution to the vapor phase. Conversely, an increase of the partial pressure of oxygen causes carbon dioxide to be driven off. These interactions, which we shall consider later in detail, are of the greatest importance for the biological function of hemoglobin. To explain them, we must assume that oxygen and carbon dioxide both react, directly or indirectly, with the hemoglobin molecule, and that the effect produced by the one reaction is somehow transmitted through the molecule so as to affect the site at which the other reaction takes place. We might suppose that both oxygen and carbon dioxide react at the same site on the hemoglobin molecule, and that the interaction effects observed are simply the result of competition between the two molecules. This explanation indeed holds for the system hemoglobin–oxygen–carbon monoxide; but for oxygen and carbon dioxide the evidence is clear that at least two separate reactive sites are involved, and that the state of either one, as determined by the extent to which it has undergone its characteristic reaction, strongly influences the reactivity of the other.

This is a relatively simple example of organization in a biochemical system. More complex examples are found in the systems of enzymes which catalyze the successive steps involved in the biological oxidation of organic molecules; the whole process runs smoothly and is highly coordinated, to an extent that implies some degree of organized arrangement in the enzyme systems involved—an arrangement still very imperfectly understood. Examples of biochemical organization may be multiplied indefinitely, but these may suffice here.

Later we shall discuss highly complex biochemical molecules and systems. Before becoming involved in these complexities, however, it is well to consider some of the basic simplicities of the chemical elements, and their distribution in living organisms.

Distribution of the Chemical Elements in Living Organisms

Living organisms as we know them utilize only a limited number of chemical elements. With very few exceptions the elements that have been clearly demonstrated to play a significant part in them are those with atomic numbers of 30 (zinc) or below. The major elements found in all living organisms are four, all with atomic numbers less than 9, namely, hydrogen, carbon, nitrogen, and oxygen. In the next row of the periodic table, sodium, magnesium, phosphorus, sulfur, and chlorine are elements of vital importance universally present. Passing beyond the electronic level of the completed argon shell with its 18 electrons, it is found that potassium, calcium, and iron are invariably essential constituents of living organisms though in relatively small amount. Cobalt, copper, and zinc, like iron, are almost invariably present although in relatively small traces. All these metallic elements, especially iron, play essential catalytic roles in certain enzyme systems, just as they function catalytically in many of the systems developed by the industrial chemist. Of the first thirty elements in the periodic table, from hydrogen to zinc, it is thus found that approximately half are invariably or frequently essential to the living organism. These first thirty elements are listed in Table IA, with a tabulation of their electronic structure and of their relative distribution in the universe at large, the earth's crust, and the human body. Comments on their biological significance are given in Table IB.

When we consider the other seventy elements of atomic numbers from 31 to 100, which are known today, it is remarkable that only three have been clearly recognized as important to the living organism. Two of these are halogens. Bromine (atomic number 35) is found only in traces in most organisms though in larger amounts in certain marine organisms. Iodine (atomic number 53), although present only in small amounts, is of fundamental and apparently universal importance. This one element of relatively high atomic number stands out in lonely eminence as a substance essential to life; all its neighbors in the periodic table appear to be either useless or harmful to living organisms, although rubidium has been found by D. Bertrand to be concentrated in certain fungi. In addition to these two halogens, molybdenum (atomic number 42) has very recently been reported as an essential constituent of certain oxidative enzymes, the molybdoflavoproteins. It is interesting that chromium (atomic number 24), which stands directly below molybdenum in subgroup VIA of the periodic table, has never yet been shown to play a role in any functioning biochemical system, in spite of the important technological use of chromium compounds in the tanning of leather.

In speaking of many of the elements as not essential to life, it must be remembered that we are speaking largely on the basis of ignorance. It is

TABLE IA
ATOMIC STRUCTURES AND DISTRIBUTION IN NATURE OF THE FIRST THIRTY
CHEMICAL ELEMENTS

Symbol	Atomic number	Electronic structure							Relative abundance		
									Universe (atoms per 10,000 atoms silicon)	Relative amount in earth's crust (atom %)	Human body (relative number of atoms)
		1	2	3	4	5	6	7			
		s	s	p	s	p	d	s			
H	1	1							3.5×10^8		9200
He	2	2							3.5×10^7		
Li	3	2	1						1		
Be	4	2	2						0.19		
B	5	2	2	1					0.23		
C	6	2	2	2					80,000		1620
N	7	2	2	3					160,000		370
O	8	2	2	4					220,000	62.55	3900
F	9	2	2	5					90		
Ne	10	2	2	6					*		
Na	11	2	2	6	1				462	2.64	11.3
Mg	12	2	2	6	2				8870	1.84	1.65
Al	13	2	2	6	2	1			882	6.47	
Si	14	2	2	6	2	2			10,000	21.22	0.14
P	15	2	2	6	2	3			130		20.3
S	16	2	2	6	2	4			3500		20
Cl	17	2	2	6	2	5			170		5
A	18	2	2	6	2	6			*		
K	19	2	2	6	2	6	1		69.3	1.42	5.6
Ca	20	2	2	6	2	6	2		670	1.94	34.4
Sc	21	2	2	6	2	6	1	2	0.18		
Ti	22	2	2	6	2	6	2	2	26.0		
V	23	2	2	6	2	6	3	2	2.5		
Cr	24	2	2	6	2	6	5	1	95		
Mn	25	2	2	6	2	6	5	2	77		0.002
Fe	26	2	2	6	2	6	6	2	18,300	1.92	0.09
Co	27	2	2	6	2	6	7	2	99		
Ni	28	2	2	6	2	6	8	2	1340		
Cu	29	2	2	6	2	6	10	1	4.6		0.0063
Zn	30	2	2	6	2	6	10	2	1.6		0.038

Data for distribution of the elements in the universe and in the earth's crust from Mason (1952), pp. 21, 42.

* Estimates are relatively uncertain.

TABLE IB
BIOLOGICAL SIGNIFICANCE OF THE FIRST THIRTY CHEMICAL ELEMENTS

Atomic number	Symbol	Comments
1	H	Universal and indispensable.
2	He	Inert.
3	Li	Apparently insignificant.
4	Be	Poisonous.
5	B	An essential constituent of plants; present also in traces in animals. Biological function little understood.
6	C	The essential basis of all forms of living matter.
7	N	Universal and indispensable.
8	O	Universal and indispensable.
9	F	An important minor constituent of vertebrate teeth; poisonous in excess.
10	Ne	Inert.
11	Na	Indispensable in animals; the predominant ionic constituent of blood plasma. Its significance in plants is more obscure than in animals.
12	Mg	Indispensable; an essential cofactor for many enzymes.
13	Al	Apparently of minor importance.
14	Si	Major element in the earth's crust; a structural component of diatoms and other forms. Probably of importance to many flowering plants.
15	P	Essential for biochemical processes. Energy transfer reactions in biochemistry generally involve phosphate esters. All nucleic acids and some proteins contain phosphorus.
16	S	Essential constituent of proteins, of coenzyme A, glutathione, and many other important biological compounds.
17	Cl	Almost universally present in animals; less important in plants.
18	A	Inert.
19	K	Major ionic constituent of many tissues; fundamentally important in such processes as nerve conduction.
20	Ca	Also of first importance; frequently significant as cofactor in enzyme reactions. Importance in plants less clear than in animals.
21	Sc	Apparently unimportant.
22	Ti	Apparently unimportant.
23	V	Essential constituent of hemovanadin, a metabolically active pigment in ascidians. Probably essential in lower plants.
24	Cr	Apparently unimportant.
25	Mn	An important catalytic factor in many biochemical reactions.

TABLE IB (*Continued*)

Atomic number	Symbol	Comments
26	Fe	Of pre-eminent importance as a catalytic factor in biological oxidations—in hemoproteins such as cytochrome oxidase, the cytochromes, catalases, and peroxidases—and for oxygen transport in hemoglobin.
27	Co	Essential constituent of vitamin B ₁₂ ; required only in traces but important.
28	Ni	Apparently of minor importance in biology.
29	Cu	An essential constituent of important oxidative enzymes and of the oxygen transporting proteins of many marine animals (hemocyanins).
30	Zn	An essential constituent of the enzymes carbonic anhydrase, carboxypeptidase, alcohol dehydrogenase, and of some other important biochemical systems.

only relatively recently, for instance, that the important role of molybdenum in living organisms has been appreciated; and biological functions for other elements, as yet unrecognized, may be discovered in future. In any case, many of the elements which we have classed as nonessential are often present in organisms, sometimes in considerable quantity. Thus although calcium is essential, and strontium and barium are not known to play any essential role, considerable amounts of the latter elements are often found in plants. For instance, Bowen and Dymond (1955) found large amounts of strontium, up to 2.6% of the total dry weight, in certain plants growing in English soils which were rich in strontium. Indeed a number of plants studied by them were found to absorb strontium preferentially, rather than calcium, from the surrounding soil. Appreciable amounts of barium were also taken up, in some cases as much as 0.7% of the dry weight of the plant, from soils rich in barium, although barium was never absorbed in preference to calcium. Such observations gain added interest from the fact that radioactive strontium (Sr⁹⁰) has become widely distributed over the earth, in small amounts, in the debris from the explosions of nuclear bombs. This radioisotope, with a half-life of approximately 27 years, is carried high into the stratosphere by the explosion of a thermonuclear bomb; it descends slowly into the lower atmosphere (the troposphere) over a period of many years and is carried down by rain onto the surface of the earth, where it is absorbed by plants, and thence passes into the bodies of animals which eat the plants (Libby, 1956). Like calcium it is concentrated preferentially in the bones and milk of these animals. In sufficient doses it produces bone tumors and

other tumors in animals and presumably in man. The tendency to absorb strontium from the soil is most marked when the soil is poor in calcium, as for example in some regions in Wales. The extent of the hazard to human and animal life which this situation presents has been sharply debated; but, important as the subject is, we shall not attempt to pursue it further here.

The distribution of the chemical elements in plants and animals is profoundly different from that found in their surroundings. In order to grow and develop, they must concentrate some elements from their environment and reject others. Since these organisms emerged from the ocean and live on the surface of the earth or within its waters, it will be profitable to consider here the conditions of their emergence and the distribution of the chemical elements in the surrounding universe and in the crust of the earth and in the oceans in which life arose.

The Relative Abundance of the Chemical Elements in the Universe

A large amount of information has now accumulated regarding the relative proportions in which the different chemical elements exist in the universe. Much of this is derived from spectroscopic studies of the sun and stars; chemical analyses of meteorites have also furnished important information. There are, of course, considerable variations in the composition of different types of stars, but for our purposes in this discussion these are relatively minor. Detailed tabulations will be found in several of the references given at the end of this chapter. The results are most conveniently summarized in the form of a diagram (Fig. 1). In this figure, the logarithm of the relative abundance of each element is plotted against its atomic number. The elements of odd and even atomic numbers are indicated by different types of symbols. Several notable features are immediately apparent from the diagram.

1. On the whole, the elements of low atomic number are much the most abundant, hydrogen and helium far more so than any of the others. Lithium, beryllium, and boron are much lower in abundance than any of their neighbors, but apart from this low region one may say in a general way that there is a gradual decline in relative abundance with increasing atomic number until values of about 40 are reached. Above atomic number 40, the relative amounts are distributed somewhat more evenly about a logarithmic value of -2 on the indicated scale.

2. Elements of even atomic number are almost invariably more abundant than the adjoining elements of odd atomic number. This relation was long ago pointed out, by Oddo in 1914 and by Harkins in 1917.

3. The value for iron (atomic number 26) is very much higher than that of any of the neighboring elements. Iron is at least a thousand times

as abundant as would be expected from an average curve drawn through most of its neighbors. Nickel (atomic number 28) is also well above the general level of such a curve, though not so high as iron.

4. Only eleven elements may be said to show really high abundance—H, He, C, N, O, Ne, Mg, Si, S, Fe, and Ni.

Undoubtedly these estimates will require some revision in future. Nevertheless, it is a striking fact that the relative proportions of the different elements are found not to differ very greatly throughout the entire known universe. This remarkable fact has naturally given rise to much

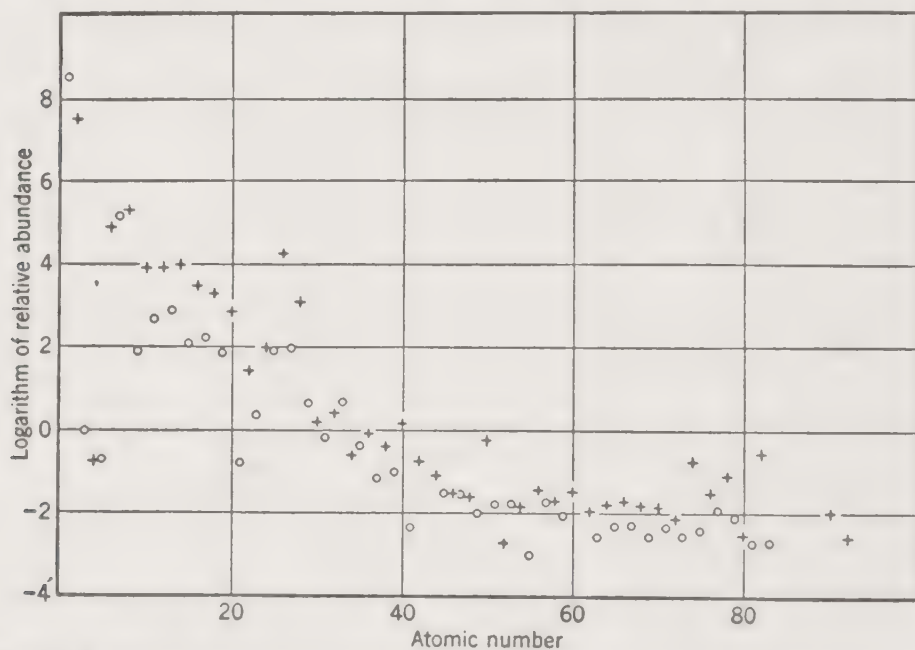


FIG. 1. Relative cosmic abundance of the elements as a function of atomic number. Elements of odd atomic number are denoted by circles, those of even atomic number by crosses. (From Mason (1952), p. 22.)

discussion as to the origin of the elements in the early stages of the development of the universe. Attempts have been made to calculate the conditions which would give rise to the observed distributions, but this is not the place to enter into these cosmological discussions. Those who are interested can learn more about them by consulting such writings as those of Alpher and Herman which are listed in the bibliography.

The Structure of the Earth and Its Crust

The mean density of the earth is approximately 5.5, whereas that of the rocks on the earth's surface is only about 2.8. Therefore, at least a large part of the interior must have a density greater than 5.5 to account for that of the whole earth. Studies of earthquake waves, and other lines

of evidence, have indicated that presumably there is a central core composed chiefly of iron, with some nickel, beginning approximately 2900 kilometers below the surface. The assumed composition of this core is closely similar to the observed composition of the iron-containing meteorites. Around this core is a mantle containing much less iron and large proportions of certain silicate rocks. Finally, near the surface is a relatively thin crust, also made up predominantly of silicates. The average composition of the crust is determined primarily by that of the igneous rocks, since the total amount of the sedimentary and metamorphic rocks is small compared to the bulk of the igneous rocks. In Table IA, an estimate is given of the average amounts of some of the most important elements in the earth's crust, based on studies by F. W. Clarke and H. S. Washington, and by V. M. Goldschmidt and others. It will be seen that the distribution of the elements in this crust is strikingly different in many respects from that of the universe as a whole. Hydrogen and helium are present in much smaller amounts, having been largely lost from the earth because of their volatility. Iron, although fairly abundant, is relatively much less so than in the universe as a whole; presumably most of the earth's iron has been concentrated in the core, leaving comparatively small amounts in the crust. The dominant elements are oxygen and silicon. Carbon, which more than any other element is characteristic of living organisms, is present in relatively minor amounts, being exceeded by at least a dozen other elements. Life has emerged and persisted in surroundings composed primarily of silicate rocks, with large amounts of water.

The structure of the earth's crust, as we know it, is the product of a long series of processes, involving the gradual selective crystallization of the most insoluble material from molten rock masses (magma) on cooling, and erosion and sedimentation processes due primarily to the action of water, continuing over vast periods of time. The details of these processes offer a fascinating study in physical chemistry. Many of the principles involved have now been worked out, owing particularly to the searching studies of V. M. Goldschmidt and his school—see, for instance, the books by Mason (1952) and by Rankama and Sahama (1950)—although most of the problem still remains to be explored. It would take us too far afield to attempt a discussion of the principles involved, but one interesting consequence of these natural processes is worth pointing out. It may be seen from Table IA that the cosmic abundance of magnesium is nearly twenty times that of sodium, and that of calcium is about ten times that of potassium. This accords with the general rule that elements of even atomic number are of higher cosmic abundance than their odd-numbered neighbors. Yet in the earth's crust all these four elements are of nearly

equal abundance—a fact which can be explained in terms of the physical chemistry of the formation of the earth’s crust, and which is certainly of major importance in determining the nature of the environment in which living organisms arose and evolved.

The mean elementary composition of living matter on the surface of the earth has been calculated approximately by the Russian geochemist Vinogradov. His data are largely derived from the wet wood of forest trees, for these make up about 80% of the living matter on land. These

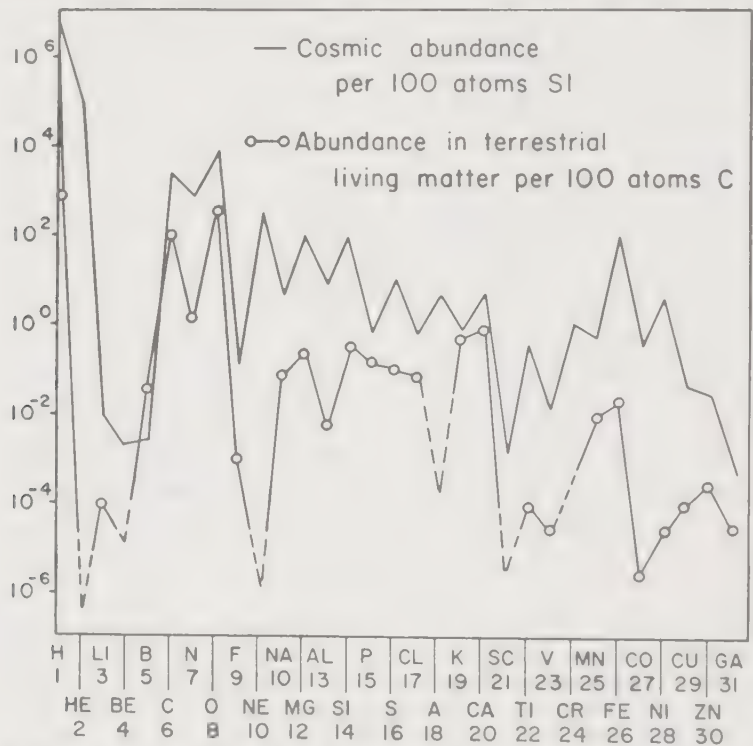


FIG. 2. Relative abundances of the first 31 elements in the cosmos and in living organisms on the earth’s surface. (From G. E. Hutchinson, 1943.)

data have been supplemented, however, by a number of analyses of other kinds of organisms. The data for the major elements H, O, C, N, K, S, P, Ca, and Mg in living organisms are certainly satisfactory. Those for the minor elements are certainly much less reliable. A revised estimate of the total distribution of these elements in living organisms was presented by G. E. Hutchinson (1943), and the results are summarized in Fig. 2, which shows the comparison between the relative abundances of the first thirty-one elements in the cosmos and in terrestrial living matter. The general parallelism of the two curves is strikingly apparent. The upper curve in Fig. 2 for cosmic abundance of the elements, as given by Hutchinson in 1943, would require some revisions in the light of knowledge

accumulated since. Such corrections, however, would not be great enough to alter the form of the curve in any very striking manner.

It may be noted that the abundance of carbon is high in both curves of Fig. 2, although a similar plot of the abundance of the elements in the earth's lithosphere would place carbon much lower on the scale. If we leave out of consideration the inert gases which, of course, cannot be retained by living organisms in significant amounts, the main biological elements are the cosmically common elements, especially H, O, C, and N. Silicon and aluminum, which are abundant on the surface of the earth, are much less prominent in the curves of Fig. 2.

Early History of the Earth

It is generally believed that the sun, the earth, and the other planets were all formed from a common pool of primitive material. It had been widely held in earlier years that the earth was formed by condensation from a very hot system composed of incandescent gases. On the other hand, the prevailing view of many, perhaps most, recent workers is that it was formed by gradual accumulation of particles from a cloud of cosmic dust at relatively low temperatures. On either basis, it would be possible to explain the relatively small amounts of hydrogen and helium in the earth as compared with the sun or stars. If the earth was formed by condensation of a mass of hot incandescent material, the lighter gases would naturally tend to escape from the earth's gravitational field during cooling. On the other hand, if the earth grew from a gradual condensation of small solid particles at rather low temperatures, the very light gaseous elements would never have taken part in the accumulation at all. The fact that the heavy inert gases krypton and xenon are also of very low abundance in the earth as compared to their distribution in the universe at large would favor the hypothesis of a gradual accretion at low temperature. If they had condensed with other elements from a hot gaseous system, their mass is high enough so that gravitational attraction would have held them on the earth.

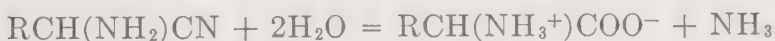
Even if the earth developed by slow accretion of particles from a cold cosmic dust cloud, it seems almost certain that it must have been very hot and molten at an early stage in its history. The assumption of such a molten state appears to be required by the marked stratification of density at different levels with a high concentration of iron and nickel in the inner core. The production of such heat, however, appears perfectly possible; the kinetic energy of the particles coming together to form the earth and the compression by gravitational attraction would generate great heat. Much heat was undoubtedly produced also by radioactive disintegrations which still continue. On the accretion hypothesis the primitive

atmosphere probably contained large amounts of water, ammonia, hydrogen, methane, and other volatile compounds that condense only at relatively low temperatures. It appears almost certain that the concentration of oxygen in the primitive atmosphere was very small, perhaps vanishingly small, and that oxygen for the most part has been produced subsequently during the earth's later history. The biological implications of these facts with respect to the possible conditions under which life originated are profound. The hypothetical earlier oxygen-free atmosphere would have favored reduction rather than oxidation of any organic compounds that might have arisen by chemical or photochemical action. Thus many organic compounds which would have been permitted to accumulate under such conditions would have been destroyed over long periods of time if the atmosphere had then contained large amounts of oxygen. It is quite possible, as was apparently first suggested by Oparin (1938, 1953), that the accumulation of organic compounds in the absence of oxygen was a necessary preliminary condition for the origin of life. The later developments of much more complex organisms from the very primitive systems which may have been the first things that could have been called living would, of course, have been impossible in the absence of oxygen and of the free energy derivable from oxidative reactions.

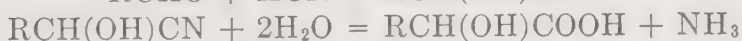
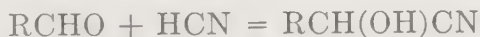
A primitive atmosphere which lacked oxygen would also have lacked ozone. The absence of the latter is important, since there is clear evidence that much of the absorption of the short-wave ultraviolet radiations from the sun—roughly in the region between 2000 and 3000 Å—is now effected by the ozone layer in the upper atmosphere. In the earlier stages of the earth's development, therefore, before this ozone layer had developed, the material on the surface of the earth was presumably exposed to an intense ultraviolet radiation, far greater than anything which occurs at the present time. This must have given rise to many photochemical reactions which cannot occur today. Moreover, there is every reason to suppose that powerful electric storms with lightning discharges occurred much as they do today. The action of short-wave ultraviolet light on a system of gases comparable to the presumed reducing atmosphere of the primitive earth has not yet been experimentally studied; such a study would be of great interest, but would also be technically difficult for various reasons. A remarkable series of experiments, however, has been reported by S. L. Miller (1955) in H. C. Urey's laboratory at the University of Chicago, making use of the action of a spark coil or of a silent electrical discharge on a gas phase containing ammonia, water, methane, and hydrogen, all of which should have been present in the atmosphere of the early earth according to the hypotheses developed by Urey and others. In Miller's experiments the mixed gases were circulated through the ap-

paratus and were passed continuously through a liquid aqueous phase, so that the involatile products of the reactions which occurred were condensed and trapped in the liquid phase. The temperature of the vapor phase was generally in the range 70° to 80°. Either the high-frequency spark discharge, at about 60,000 volts, or a silent discharge yielded considerable quantities of organic compounds in the course of a week's run, but the yield was considerably greater with the spark coil. In either case, a remarkable variety of organic compounds was isolated at the end of the experiment. These included considerable amounts of glycine, alanine, β -alanine, sarcosine (N-methylglycine), α -aminobutyric acid, α -aminoisobutyric acid, and probably some aspartic acid also. A large number of other amino acids were present in small amount, including several that could not be identified. Among other compounds formed in significant amount were formic, acetic, propionic, glycolic, and lactic acids. Other polyhydroxy compounds of unknown composition also appeared to be present. Hydrocyanic acids and several aldehydes were clearly direct products of the discharge; it seems possible that the synthesis of the hydroxy and amino acids may be through the hydroxy and amino nitriles in the solution by the reactions:

1. For the amino acids:



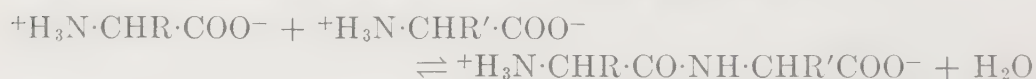
2. For the hydroxy acids:



The exact mechanism of the observed synthesis is far from being established, and Miller's experiments in any case represent only the beginning of research in this field. They provide decisive evidence, however, that many of the amino acids and hydroxy acids which we know to play an essential part in the structure and activity of living organisms can be produced readily and in considerable amounts under conditions closely simulating those of the hypothetical primeval atmosphere of the earth. The amino acids are extremely involatile, and once deposited in the primeval ocean they would have remained there, available for further reactions. Likewise the hydroxy acids, though much more volatile than the amino acids, are still so involatile that they also would have remained in the liquid phase.

It thus seems reasonable to believe that the action of electrical discharge or of ultraviolet light, or both, in the early stages of the earth's

history, acting over a period of many millions or even hundreds of millions of years, would have led in the course of time to the formation of a great variety of organic compounds, in large amounts in the primitive ocean. The further reactions that these compounds might have undergone are still largely a matter of speculation. For instance it has recently been suggested by H. F. Blum (1955) that the condensation of amino acids to form polypeptides could have occurred in pools of water derived from the primeval ocean, as a result of early geological changes of the earth's surface. Evaporation of most of the water in such a pool would have produced an extremely concentrated solution in which the conditions would have been relatively favorable for the spontaneous polymerization of amino acids into long peptide chains, by reactions of the type:



We shall discuss such reactions further in Chapter 3 and in later chapters. It is obvious that, since water is produced in such a reaction, the evaporation of water from the system would shift the equilibrium to the right and favor the formation of peptide ($-\text{CO}\cdot\text{NH}-$) linkages, although, as we shall see, the equilibrium conditions in a solution containing considerable amounts of water are not favorable to the condensation process involved in the formation of peptide bonds. Such reactions, however, in so far as they did occur, must have taken place in the presence of a large number of other organic substances capable of undergoing many reactions. These polypeptides and other constituents would have tended to be absorbed upon particles of clay, resulting from erosion—a possibility first suggested by Bernal (1951). As Blum (1955) remarks, "The polypeptides absorbed on, say, the surface of clays, might have been in a very good position to enter upon a catalytic role. These polypeptides would not have been the compact proteins, nor the efficient enzyme catalysts found in modern living organisms, but they could well have been their early ancestors."

All this is naturally in large measure speculative and inconclusive. Miller's experiments, however, have in a striking way demonstrated the possibility of experimental study of systems simulating the possible conditions on the early earth—conditions which have long since vanished—and of drawing reasonable inferences concerning the development of complex systems of organic and inorganic compounds which were the necessary prelude to life as we know it.

How long such early conditions could have lasted, in the presence of a reducing atmosphere, is difficult to say. It seems probable, however, that the introduction of oxygen into the atmosphere began at a relatively early stage of geological history. How its formation began is debatable. After

some oxygen was present and plant life with its capacity for photosynthesis had arisen, it is almost certain that the oxygen produced by photosynthesis came to account for a very large part of the oxygen in the atmosphere. Since some oxygen must have been present, however, before plants with the capacity for photosynthesis could develop, it is clear that other reactions leading to the production of oxygen must have occurred. It has been suggested that this arose from the thermal dissociation of water vapor when the temperature of the earth's surface was sufficiently high—say about 1500° —or else by photochemical dissociation of water vapor by solar radiation in the upper atmosphere. The change from a reducing to an oxidizing atmosphere must have been one of the great crucial transitions in the history of the earth.

The conceptions of the early atmosphere here discussed are necessarily speculative and open to question. The geologist W. W. Rubey (1955) has made a searching inquiry into the fundamental equilibria involving water, methane, ammonia, nitrogen, chlorine, hydrochloric acid, sulfur, hydrogen, hydrogen sulfide, and carbon dioxide under conditions that might be anticipated in the earth's early history. His calculations, assuming equilibrium to be attained with respect to all the reactions considered by him, early in the course of geological time, give very different results from those of Urey for the composition of the early atmosphere. Rubey's estimate for such an atmosphere is 91% carbon dioxide, 6% nitrogen, 2% hydrogen sulfide, with minor amounts of other constituents. Rubey admits the possibility that an atmosphere such as that postulated by Urey may have existed in the very early stages of the earth's development, but he believes that before long it must have been superseded by one of a very different sort. The assumption that equilibrium was attained with respect to all possible reactions remains uncertain, however, as Rubey is careful to point out. It is quite likely, in any case, that large regions on and just under the earth's surface may have existed under reducing conditions, early in the earth's history. Rubey suggests that there were "local environments in the ocean and in mud pools where free oxygen was absent; and that it was in such local reducing environments that the first organisms may have come into existence" (1955, pp. 648–9).

Abelson (1956) has taken up the inquiry again, employing various combinations of gases and subjecting them to electric discharges by a technique similar to that of Miller. Abelson has, however, concluded that the atmosphere proposed by Rubey would be modified under the influence of ultraviolet radiation. A series of reactions would be expected to occur, involving the reducing alkaline crust of the earth, which would tend to alter the composition of the atmosphere. For instance, H_2O can be dissociated by short-wave ultraviolet light to give the free radicals

H and OH, both of which are very reactive. They may recombine or may undergo various reactions such as the formation of H_2 from $H + H$ or H_2O_2 from $OH + OH$. One effect of these reactions would be to liberate hydrogen into the atmosphere and make it more reducing. Therefore Abelson, in repeating Miller's experiments with various combinations of gases, added hydrogen as one of the components of the system. He studied the mixtures $CO_2-N_2-H_2-H_2O$, $CO-N_2-H_2-H_2O$, and $CO_2-NH_3-H_2-H_2O$, and in every case observed the formation of amino acids such as alanine, β -alanine, glycine, and sarcosine. It is thus apparent that a considerable number of possible hypotheses about the composition of the primitive atmosphere and the earth would all be compatible with conditions which could give rise to such substances as amino acids by the action of ultraviolet light or of electrical storms.

There is, of course, a huge gap between the early formation of a large variety of organic compounds, in the presence of the postulated reducing atmosphere, and the appearance of organized metabolic systems, capable of reproduction, which we could definitely recognize as alive. The early geological record has been far too completely obscured to permit much hope that we shall ever recover any direct evidence of this crucially important transition period which led to the production of the first organisms which might have been recognizable as such. Probably the transition period was very long, and involved many intermediate forms which we should scarcely recognize as living organisms at all, by usual standards. No doubt many inefficient processes of metabolism developed and were discarded. Thoughtful, but necessarily inconclusive, discussions of many of the problems involved have been given by a number of authors; we may mention particularly Oparin (1938), Bernal (1951), and Blum (1951, 1955).

Such compounds as amino acids and porphyrins, which are essential constituents of living organisms as we know them today, were certainly in existence hundreds of millions of years ago. Thus Abelson (1953-54) has identified the amino acids glycine, alanine, leucine, valine, aspartic acid, and glutamic acid in fossils of the fish *Dinichthys*, dating from the Devonian period about 360,000,000 years ago, and in a number of other fossils more than 100,000,000 years old. Conclusive proof that these amino acids were a part of the original organism, and were not trapped later within the structure, is difficult to obtain; but Abelson has produced fairly strong evidence for this conclusion. Moreover he has shown that these amino acids are highly stable; from the rate of destruction at high temperatures he has estimated—by a calculation which involves a considerable extrapolation—that alanine, for instance, is stable enough to last for more than 10^9 years at ordinary temperatures.

The porphyrins, which are organic structures of great stability, form complexes with magnesium to produce chlorophyll, and with iron to form the various heme proteins—cytochrome oxidase, the cytochromes, catalase, peroxidase, and the hemoglobins—which in one form or another are found almost universally in living organisms. Porphyrins have been identified in shales, asphalts, and petroleum dating back to the Paleozoic period (Mason, 1952, p. 210); hence they are of very great antiquity. Nevertheless the earliest organisms from which such compounds were derived were probably relatively advanced; the nature of the very earliest organisms remains inaccessible to any direct evidence.

The Ocean

There seems little doubt that the most primitive forms of life must have developed in the ocean. The extraordinary properties of water and their significance for biology will be discussed in detail in the next chapter. Here we need only say that water is indispensable to life as we know it and almost certainly indispensable to any conceivable form of life that might arise within the universe. No other chemical compound can substitute for water as far as we know.

The history of the development of the ocean is still a matter of debate. Obviously liquid water could not form on the earth's surface until this surface had cooled at least to a temperature below the critical point of water (370°). Since there is no way of estimating with any assurance even the approximate amount of water in the primeval atmosphere, the size of the primitive ocean which first appeared on cooling must still be a matter of speculation. Some have suggested that the oceans of the earth had approximately their present volume even at a very early stage. Others have held that the primitive ocean was very small, and that its volume has grown gradually by the addition of water arising from the interior of the earth—some water being liberated from molten igneous rock on cooling, some through volcanic activity and in other ways.

Leaving aside for the moment the question of the origin and development of the ocean, let us consider the facts concerning the state of the oceans today. They cover an area of $370,000,000 \text{ km}^2$, or nearly 71% of the earth's surface. If the mean depth is taken as 3100 meters, the volume of the oceanic waters is $1370 \times 10^6 \text{ km}^3$. With a mean density of 1.03, this gives slightly more than 1400×10^{21} grams for the total mass of the oceans. The total amount of fresh water and of continental ice is trivial by comparison.

Sea water, of course, contains numerous ions, of which sodium and chloride are the most abundant, and various other constituents. The

TABLE II
IONIC COMPOSITION OF SEA WATER AND OF HUMAN BLOOD PLASMA
(In millimoles and milliequivalents per liter)

Anions and weak acids	Sea water*			Blood plasma †, ‡		
	mM/l	meq/l	moles 100 moles Na ⁺	mM/l	meq/l	moles 100 moles Na ⁺
Cl ⁻	561	561	117	102	102	74
SO ₄ ⁻⁻	29	58	6	0.5	1.0	0.35
H ₂ PO ₄ ⁻	—	—	—	2	2	1.5
HPO ₄ ⁻⁻	—	—	—	1.2	2.4	0.9
Br ⁻	0.85	0.85	0.2	—	—	—
F ⁻	0.07	0.07	—	—	—	—
CO ₂ + H ₂ CO ₃	0.01	—	—	1.3	1.3	1
HCO ₃ ⁻	1.90	1.90	0.4	27	27	20
CO ₃ ⁻⁻	0.20	0.40	0.04	—	—	—
H ₃ BO ₃	0.35	—	0.07	—	—	—
H ₂ BO ₃ ⁻	0.08	0.08	—	—	—	—
Total anions	593.1	622.3		134	135.7	
Cations						
Na ⁺	481	481	100	138	138	100
Mg ⁺⁺	55	110	11.5	0.85	1.7	0.6
K ⁺	10.2	10.2	2.2	4.2	4.2	3.0
Ca ⁺⁺	10.5	21.0	2.2	2.6	5.2	1.9
Sr ⁺⁺	0.15	0.3	—	—	—	—
Total cations	556.9	622.5		145.7	149.1	

* Data taken with slight modifications from Sverdrup, Johnson, and Fleming (1942); and from Rubey (1951).

† Data from "Standard Values in Blood" (E. C. Albritton, ed.), U. S. Air Force, Dayton, Ohio, 1951.

‡ Certain anions of blood plasma are not listed here. The plasma proteins contribute approximately 14 meq/l to the anionic net charge. Certain organic anions, such as lactate and pyruvate, also make a small contribution.

amount of the major constituents of normal sea water is given in Table II. The distribution of the relative proportions of the ions present in sea water has profoundly impressed many students of biology. In particular, A. B. Macallum pointed out many years ago that the relative proportions of sodium, potassium, calcium, and chloride ion were rather similar in the ocean and in the blood plasma of mammals at the present day. The

absolute amounts of all these ions, however, are markedly greater in the ocean than in blood plasma, the ratio of ocean concentration to plasma concentration being somewhat greater than 3 for sodium ion and somewhat greater than 4 for chloride ion. Macallum was led by consideration of these relations to propose the remarkable hypothesis that the present composition of the blood plasma—the “internal environment” of the higher animals—was a direct reflection of the composition of the primitive ocean from which their ancestors had emerged in earlier ages. He supposed that the gradual influx of elements into the ocean through inflow of rivers and in other ways had raised concentrations to their present level, while maintaining essentially the same relative levels that had prevailed at the time long past when our primitive ancestors emerged from the sea. Recent data for the ionic composition of human blood plasma—which is in this respect fairly typical of the plasma of most animals—are given in Table II, for comparison with sea water.

Fascinating as Macallum's hypothesis is, it gives rise to grave difficulties and must probably be rejected. It will immediately be seen by inspection of Table II that the relative proportions of different ions in plasma and in the ocean today are in some instances markedly dissimilar. For instance, calcium ion is approximately four times as abundant in the ocean as in blood plasma, whereas magnesium ion is about sixty times as abundant. This in itself is perhaps not a grave difficulty; one might propose various special hypotheses to explain why magnesium ion might have increased in concentration more rapidly than calcium in the later stages in the development of the ocean. However, a thoughtful and detailed consideration of the chemical evolution of the ocean by E. J. Conway has led to results which appear to rule out Macallum's hypothesis entirely.

Before considering Conway's views, it will be helpful first to consider how far dissolved constituents found in the ocean today can have accumulated there by the weathering and disintegration of igneous rocks during the course of geological time. It has been estimated by the eminent Norwegian geochemist V. M. Goldschmidt that approximately 600 grams of rock has been weathered for each kilogram of water found in the ocean today. This rock may, therefore, be regarded as the potential source of much of the dissolved material found in the ocean. Most of it has passed down by means of the runoff of the rivers. Most of this rock, of course, represents insoluble matter; only a relatively small part has dissolved, and of the soluble constituents a considerable fraction has later undergone chemical reactions within the ocean to form insoluble compounds. On the basis of the known average composition of igneous rocks and of the material from them which can potentially pass into sea water, Gold-

schmidt drew up a table comparing the total amount of the various elements supplied to sea water with the amounts actually present in the sea water today. This table was revised and improved by Rankama and Sahama and is given here in abbreviated form as Table III, for some of the elements of particular interest to the biochemist. It will be seen immediately that the percentage of various elements remaining in solution varies enormously. Over 60% of the total sodium supplied to sea water, according to these estimates, still remains in solution as sodium ion, but only 2.4% of the potassium and 1.8% of the calcium still remain in solution in ionic form.

The most striking feature of Table III, however, is the demonstration that certain elements are present in sea water to an extent far greater than the weathering of rock can possibly account for. This is true notably of boron, sulfur, chlorine, bromine, and iodine—in particular for chloride and bromide ions. The concentration of chloride in the water of rivers is small compared with that of sodium, calcium, or magnesium ions. Moreover the concentration of chloride ion in inland surface waters is found to fall off rapidly with increasing distance of these waters from the sea. The matter has been analyzed by Conway in great detail; he and others have drawn the conclusion that most of the chloride in surface waters is carried by winds from sea to land in the form of fine particles of sodium chloride and other salts which settle on the earth or in its surface waters. Therefore, much the greater part of the chloride ion found in river water must actually have come from the sea, to which it is then carried back in a cyclic process. Thus, although sodium and chloride ions are the most important ions of sea water, their origins must be quite different. Most—indeed nearly all—of the sodium presumably comes from the weathering of rock and the transport of sodium ion in the form of soluble salts by rivers to the ocean. The origin of the net accumulation of chloride ion in the ocean is more dubious. Much of it has probably been derived from the gases and liquids released by volcanic action, both from surface volcanos and from those under the sea. It has also been suggested by some investigators that the primeval atmosphere contained large amounts of hydrogen chloride which condensed to give a primitive ocean of high hydrochloric acid content. If this did occur, the acidity of this earlier ocean was gradually neutralized by the basic salts arising from the weathering of rock and carried by rivers down to the sea. It seems probable, however, that most of the chloride ion—the same presumably being true of bromide and iodide—has been added gradually to the ocean over a long period of time by volcanic action.

Conway has attempted to calculate the rate at which the content of sodium ion in the ocean has increased gradually throughout geological

TABLE III
GEOCHEMICAL BALANCE OF SOME ELEMENTS IN SEA WATER

Element	Total amount supplied to sea water (g/ton)	Amount present in sea water (g/ton)	Percentage of amount supplied remaining in solution
Li	39	0.1	0.3
B	1.8	4.6	256
C	192	28	14.6
N	27.78	0.7	2.5
F	540	1.4	0.3
Na	16,980	10,561	62
Mg	12,540	1,272	10
Al	48,780	1.9	0.004
Si	166,320	4	0.002
P	708	0.1	0.01
S	312	884	283
Cl	188.4	18,980	10,074
K	15,540	380	2.4
Ca	21,780	400	1.8
Sc	3	0.00004	0.001
V	90	0.0003	0.0003
Mn	600	0.01	0.002
Fe	30,000	0.02	0.00007
Co	13.8	0.0001	0.0007
Ni	48	0.0005	0.001
Cu	42	0.011	0.03
Zn	79.2	0.014	0.02
As	3	0.024	0.8
Se	0.054	0.004	7.4
Br	0.972	65	6,687
Rb	186	0.2	0.1
Sr	180	13	7.2
Mo	9	0.0007	0.008
Ag	0.06	0.0003	0.5
Sn	24	0.003	0.01
I	0.18	0.05	28
Cs	4.2	0.002	0.05
Ba	150	0.05	0.03

This is an abbreviated form of a table by Rankama and Sahama (1950), p. 195.

time. Any such calculation, of course, involves a number of assumptions which are open to question, but those which Conway has chosen appear reasonable at least as a rough approximation. Essentially he has attempted to correlate the progressive increase in the content of sodium and related ions in the ocean with the weathering of rock and the concomitant increase in the thickness of the various sedimentary strata which have been laid down as a result of this weathering, from the earliest time to the present. He also assumed that the rate of sodium increase in the ocean is proportional to the mean sodium content of the weathering rock. The same fundamental calculation was made on several different hypotheses—on the one hand, on the assumption that the volume of the ocean has remained relatively constant, and on the other hand, on the assumption that the volume has gradually increased throughout geological time. Likewise two extreme assumptions were considered regarding the chloride ion content of the ocean: first, that this content has remained constant throughout geological time; and second, that it has increased gradually from zero to the present level. The details of Conway's calculations are far too long to discuss here, but the conclusion on any of these hypotheses appears quite unequivocal. The origin of vertebrates is generally placed at the period of the early Ordovician sea. From any of the above hypotheses the concentration of sodium and of chloride ion in the ocean at this time must have been more than twice as great as the values found in mammalian blood plasma. Even the pre-Cambrian sea—at a far earlier period, when life was at a very primitive stage—must, according to Conway's calculations, have been rather more than twice as concentrated in sodium ions as blood plasma of the present day. Therefore, the simple picture that the electrolytes of vertebrate blood plasma today reproduce the concentration of a primeval ocean from which our ancestors emerged appears to be incompatible with the available evidence.

Although we must apparently reject Macallum's hypothesis, the importance of the ocean for the development of living organisms is not thereby diminished. As L. J. Henderson has pointed out in "The Fitness of the Environment," the ocean is like a vast thermostat, in which temperature is maintained nearly constant over large regions and considerable periods of time. Temperature fluctuations at any one place are small; thus Henderson cites data recorded at Lesina, showing a mean winter temperature of the surface ocean waters of 13.5° and a mean summer temperature of 22.0° . This variation of 8.5° between winter and summer is larger than is found in most parts of the ocean. Below the surface the fluctuations in temperature from season to season are less. As the depth increases the temperature decreases, falling to a value near 2° at great depths in mid-Atlantic. Variations of average ocean temperature with

latitude, although significant, are far smaller than the corresponding variations of land temperature with latitude. In any one given region the ocean is an environment of remarkable stability and constancy for living organisms.

Variations of salt concentration between one part of the ocean and another are also small, except in the estuaries where river and ocean water mix. The average total salt content by weight is about 3.45%. In places such as the coast of Norway, and on the east coast of North America, the value is lower—3.2 to 3.3%; in most of the North Atlantic it is 3.5 to 3.6%. The *relative* proportions of the ions and molecules in sea water remain in general very constant indeed; thus, although the absolute figures given in Table II for sea water represent a typical sample from a particular region, the ratios may be taken as highly accurate for ocean water over most parts of the globe. This relatively constant salinity is of the greatest importance in providing a stable environment for living organisms.

Sea water is also a stable medium with respect to its nearly constant alkalinity. As with the blood of the higher animals, the principal factor involved in maintaining this stability is the equilibrium between carbon dioxide and the bicarbonate and carbonate ions. Sea water is slightly more alkaline than blood—in the usual notation of *pH* values the *pH* of blood is ordinarily 7.4, that of sea water 8.2. As Henderson has well written:

The regulatory devices of our modern laboratories have not yet succeeded in rivaling the ocean. Singly, certain conditions, for example, temperature, alkalinity, and concentration, may be more accurately regulated by man, though on a small scale only; but the regulation of all such properties together is not yet possible. The only known improvement upon the ocean is the body of a higher warm-blooded animal. Here, however, the processes of organic evolution have begun with the ocean, and in several respects merely perfected existing arrangements.—L. J. Henderson, "The Fitness of the Environment," p. 186.

The Central Role of Carbon in Biology

The fact that carbon compounds occupy a key position in life as we know it is so obvious as to need no comment. Inquiring and imaginative men have sometimes raised the question whether other worlds might exist, containing forms of life in which some other element—perhaps silicon—played the central part which carbon does in ours. Such views were explicitly rejected in 1913 by L. J. Henderson who, in "The Fitness of the Environment," brought together a powerful array of data to demonstrate that the elements carbon, hydrogen, and oxygen were unique in their fitness to function in systems with the characteristics of living organisms as we know them. Although Henderson's thesis has been criticized, the weight of the evidence arising from our deeper knowledge

of the properties of matter, since his book was written, has strengthened his fundamental thesis. N. V. Sidgwick, in the quotation which heads this chapter, has held that a world in which silicon should replace carbon as the basis of life is impossible. He has offered two reasons for the unique position occupied by carbon. First, the typical state of the carbon atom in its compounds, with covalent bonds, is extremely stable. The octet of four electron pairs involves full sharing of all the electrons of the carbon atom. The octet cannot be increased beyond 8 electrons, since 4 is the maximum covalence of carbon, so that the saturated carbon atom cannot coordinate either as donor or acceptor. Silicon, on the other hand, is not limited to a covalence of 4, but, as with other elements in the second row of the periodic table, this value can rise to 6. Since chemical reactions commonly proceed through coordination, carbon must in general be very slow to react. Thus carbon compounds, even those which may be thermodynamically quite unstable, tend in practice to remain stable over long periods of time, because of the difficulty of finding a path by which reaction may take place. Indeed one of the great problems which has been solved by living organisms is to develop catalysts which make such paths smoothly available when they are needed, so that the energy obtainable from the transformation of organic compounds can be released at the time and place where it is needed within the cell. The relative inertness of most carbon compounds, except in the presence of these catalysts, at ordinary temperatures, is, of course, of the greatest value in preserving the cell from wasteful or dangerous release of energy in a random fashion.

The other important reason for the great multiplicity of carbon compounds that exist is that the energy involved in the formation of C—C bonds is not very different from that for the formation of C—H, C—O, or other bonds involving carbon. In the case of silicon, on the other hand, the formation of the Si—Si bond involves much less energy than that of the Si—H or Si—O bond. The value for those energies, as tabulated by Sidgwick, are given below.

BOND ENERGIES OF CARBON AND SILICON					
	X—X	X—H	X—O	X—Cl	
X = C	81.6	98.8	81.5	78.0	kcal
X = Si	42.5	75.1	89.3	85.8	kcal
Difference C minus Si	+39.1	+23.7	— 7.8	— 7.8	kcal

Thus nearly twice as much energy is required to break a C—C bond as to break a Si—Si bond, and this greater stability of the former is certainly a major factor in permitting the existence of organic compounds containing long chains of carbon atoms.

Among the compounds of carbon, carbon dioxide occupies a unique place—there is nothing in the least like it among the silicon compounds, and indeed no other molecule resembles it at all closely. Its volatility, the ready way in which it distributes itself almost equally between a gas phase and an aqueous solution, its capacity to undergo reversible hydration to form the weak acid H_2CO_3 —all this is of the utmost importance in the maintenance of a stable hydrogen ion concentration in the oceans and other waters of the earth, and in such biochemical systems as blood. Indeed the conclusion of L. J. Henderson that carbon dioxide is uniquely fitted among all known compounds for the function which it performs appears to be substantiated by all the evidence available today. Later, in Chapter 10, we shall discuss in detail the properties and some of the biological functions of carbon dioxide. Another simple triatomic molecule, namely water, has unique properties which make it even more important than CO_2 in the biological scheme of things. It is so important that it deserves a chapter to itself, to which we now proceed.

REFERENCES

- Abelson, P. H. (1953-54). Annual Report of the Director of the Geophysical Laboratory, Carnegie Institution of Washington, Year Book No. 53, pp. 97-101.
- Abelson, P. H. (1956). Amino acids formed in primitive atmosphere, *Science* **124**, 935. A brief preliminary report on Abelson's recent work.
- Alpher, R. A., and Herman, R. C. (1950). The theory of the origin and relative abundance distribution of the elements, *Revs. Mod. Phys.* **21**, 153.
- Baldwin, Ernest (1949). "Introduction to Comparative Biochemistry," 3rd ed., Cambridge University Press, Cambridge. Contains a valuable and illuminating discussion of the ocean in relation to living organisms. The book is highly recommended reading, but it should be noted that Macallum's hypothesis regarding the relation of the ocean to living organisms is accepted somewhat uncritically.
- Bernal, J. D. (1951). "The Physical Basis of Life," Routledge and Kegan Paul, London. A brilliant short discussion closely related to that of Oparin (see below) but going beyond Oparin in certain respects.
- Blum, H. F. (1951). "Time's Arrow and Evolution," Princeton University Press, Princeton, New Jersey. The primary thesis of the book is closely related to that of Henderson, but use is made of important modern developments not available when Henderson's book was written. Valuable discussion of energetics in relation to biology.
- Blum, H. F. (1955). Perspectives in evolution, *Am. Scientist* **43**, 595. A further discussion of some of the points discussed in Blum's book referred to above.
- Bowen, J. M., and Dymond, J. A. (1955). Strontium and barium in plants and soils, *Proc. Roy. Soc.* **B144**, 355.
- Conway, E. J. (1942). Mean geochemical data in relation to oceanic evolution, *Proc. Roy. Irish Acad.* **B48**, 119-159; (1943). The chemical evolution of the ocean, *ibid.* 161-212. These two long papers contain a brilliant and detailed discussion of oceanic chemistry and evolution. For comment and criticisms, see G. E. Hutchinson, *Am. J. Sci.* **242**, 272 (1944); and also Chapter 7 in Mason, "Geochemistry," listed below.

- Goldschmidt, V. M. (1954). "Geochemistry," edited by Alex Muir, Oxford University Press, New York. This book is by one of the great masters in the field of geochemistry, left incomplete at his death and edited by Dr. Muir.
- Henderson, L. J. (1913). "The Fitness of the Environment," The Macmillan Company, New York. A pioneer work of fundamental importance for the understanding of the biological significance of the properties of matter. The chapters on water, carbon dioxide, and the ocean are especially important in the present connection.
- Hutchinson, G. E. (1943). The biogeochemistry of aluminum and of certain related elements, *Quart. Rev. Biol.* **18**, 1, 128, 242, 331. This comprehensive review has a far wider scope than its title might suggest. It contains valuable information on a great many aspects of biogeochemistry. It is well worth reading for the perspectives it offers.
- Libby, W. F. (1956). Radioactive strontium fallout, *Proc. Natl. Acad. Sci. U. S.* **42**, 365; also *ibid.* **43**, 758.
- Macallum, A. B. (1926). The paleochemistry of the body fluids and tissues, *Physiol. Rev.* **6**, 316.
- Mason, Brian (1952). "Principles of Geochemistry," John Wiley & Sons, New York. An excellent introduction to the subject in a relatively short book. Contains valuable chapters on the abundance of the elements in the hydrosphere, atmosphere, and biosphere which are closely related to the subject matter of the present chapter.
- Miller, S. L. (1955). Production of some organic compounds under possible primitive earth conditions, *J. Am. Chem. Soc.* **77**, 2351.
- Oparin, A. I. (1938). "The Origin of Life," The Macmillan Company, New York, reprinted by Dover Publications, New York, 1953. The ideas advanced in this book have been very influential. A new edition, "The Origin of Life on the Earth" (1957), translated by Ann Synge, published by Oliver and Boyd, London, and Academic Press, New York, has appeared while this book was in press.
- Pauling, L. (1940). "The Nature of the Chemical Bond," 2nd ed., Cornell University Press, Ithaca, New York. A fundamental book for the interpretation of the properties of chemical compounds in terms of their structure.
- Rankama, K., and Sahama, T. G. (1950). "Geochemistry," University of Chicago Press, Chicago. A comprehensive, detailed treatise much more extensive than that of Mason. The general discussion is clearly written; it is also useful for reference on many special points.
- Rubey, W. W. (1951). Geologic history of sea water. *Bull. Geol. Soc. Am.* **62**, 1111. An important discussion of oceanic evolution.
- Rubey, W. W. (1955). Development of the hydrosphere and atmosphere, with special reference to probable composition of the early atmosphere, *Geol. Soc. Amer. Spec. Paper* **62**, 631-650. Reprinted from "The Crust of the Earth."
- Sidgwick, N. V. (1950). "The Chemical Elements and Their Compounds," 2 vol., Oxford University Press, New York.
- Sverdrup, H. U., Johnson, W., and Fleming, H. (1942). "The Oceans, Their Physics, Chemistry and General Biology," Prentice-Hall, New York. A comprehensive treatise of great value.
- Urey, H. C. (1952). "The Planets," Yale University Press, New Haven, Connecticut. An important discussion of the origin and evolution of the planets, including the earth. Many of the views expressed are still the subject of considerable controversy.
- Wells, A. F. (1950). "Structural Inorganic Chemistry," 2nd ed., Oxford University Press, New York. Very helpful for discussion of structural principles, with many details concerning atomic arrangements.

Chapter 2

Water and Its Biological Significance

The structures of living cells and tissues are profoundly complex; they involve the most complicated molecules known (the proteins) and a great variety of carbohydrates, fats, phospholipids, steroids, and other substances which have important and specific activities. Certainly, moreover, these structures are not distributed at random, but are arranged in a highly organized pattern within the cell. This organization is not, of course, static; it is highly mobile and is undergoing constant transformation. It is not the fixed and nearly perfect, but relatively static, type of order that is to be found in a true crystal, such as diamond, quartz, or rock salt. It is, however, closer in nature to crystalline order than to the molecular chaos found in gases, or to the relatively random arrangement of the molecules found in many simple liquids.

All the complex substances mentioned above, however, make up only a minor portion of the total weight of living tissues; the principal component is water, present in amounts far exceeding the total of all the other components. Thus 93% of the blood plasma, and about 60% of the red blood cells, is water; muscle tissue is approximately 80% water, and water makes up considerably more than half—often considerably more than three-quarters—of most other tissues. The only notable exceptions are certain relatively inert tissues, such as hair, horn, and the solid portion of bone. The spores of certain plants and bacteria have a low water content, but these spores are relatively inactive cells. When they are transformed into cells which show active metabolism and growth, the increase of water content is an essential part of the transformation. Thus water appears to be the indispensable matrix for the structural components and the activities of living cells. Because it is so universally present and so important in biology, its importance is often taken for granted and it receives little explicit discussion in many treatises on biochemistry. It is desirable to consider its properties and functions at this point in some detail.

Water is so familiar that it would appear odd to call it a substance of strange and unusual properties. Yet the chemist, on comparing it with other substances of closely related electronic structure, finds its behavior full of apparent anomalies. Consider the series methane, ammonia, water,

hydrofluoric acid, neon. The number of atoms in each molecule decreases from 5 (methane) to 1 (neon), but the aggregate charge associated with the atomic nuclei remains constant throughout the series and is equal to 10 proton units.¹ Associated with these in each case are 10 electrons in all, which make the entire structure neutral. The positive charges on the central heavy nucleus increase in this series from six to ten; the positive charges on the attached hydrogen nuclei decrease from four to zero. The physical properties of these five substances vary in a most striking fashion; we may consider for instance their melting points, boiling points, and molal heats of vaporization (Table I). The values of all these properties

TABLE I
SOME PHYSICAL PROPERTIES OF A SERIES OF ISOELECTRONIC SUBSTANCES

Substance	Melting point (°C)	Boiling point (°C)	Molal heat of vaporization (cal/mole)
CH ₄	-184	-161	2200
NH ₃	-78	-33	5550
H ₂ O	0	+100	9750
HF	-92	+19	7220
Ne	-249	-246	415

rise to a maximum for water, midway in the sequence, and are lowest for methane and especially for neon. It is illuminating to consider these properties also for corresponding sequences of substances in the higher rows of the periodic table. The results are shown in Figs. 1 and 2, taken from Pauling (1940), in which the rare gases have been omitted from the plot. For a series of compounds RH_n, if *n* is constant, the melting point, boiling point, and heat of vaporization all increase with increase in the atomic weight of R, *provided we ignore the elements in the first row of the periodic table*. In the series CH₄, SiH₄, GeH₄, SnH₄, it is found that the values for methane fall very close to the points obtained by extrapolating backward the points for the other compounds; likewise for the rare gases there is a progressive decrease, in the order Xe, Kr, A, Ne. The values for HF, NH₃, and H₂O, however, lie far above the values predicted by extrapolation from the values for their higher homologs. This is especially true of water, for which a boiling point near -100°, instead of +100°, would be predicted by back-extrapolation of the line through H₂Se and H₂S.

¹ One proton unit = 4.803×10^{-10} electrostatic unit of charge. The charge of an electron is, of course, equal to minus one proton unit, and it is the electronic charge which has generally been determined by direct experiment.

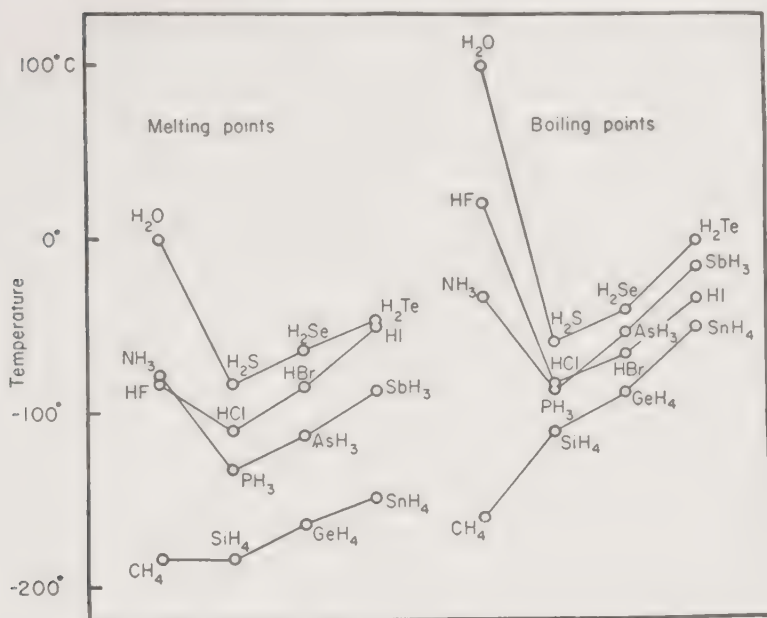


FIG. 1. Melting and boiling points of isoelectronic sequences of hydride molecules. (From Pauling, 1940.)

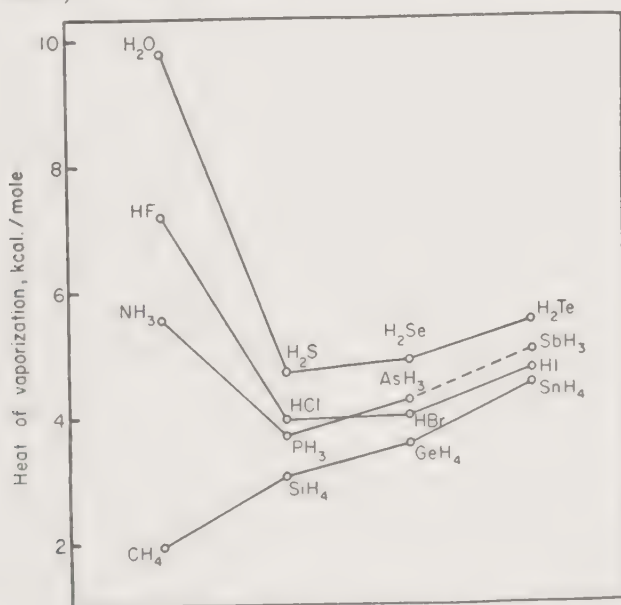


FIG. 2. Heats of vaporization of isoelectronic sequences of hydride molecules. (From Pauling, 1940.)

High melting and boiling points and high heats of vaporization are indicative of strong intermolecular forces. The heat of vaporization (or sublimation), indeed, directly measures the energy necessary to separate the molecules of a liquid (or solid) from one another. Thus, of the substances listed in Table I, methane and neon—especially the latter—are

characterized by low intermolecular forces. In the case of neon, the reason is obvious in the light of modern valence theory. The 2 inner (*K*) electrons and the 8 outer electrons form a stable system, with no tendency either to add or to lose electrons. The negative charges on the electrons are distributed symmetrically about the positive charge on the nucleus. Because of this symmetry, electrical forces between neon atoms are almost nonexistent; only the very weak forces of intermolecular attraction, known as van der Waals' forces, are present. The case of methane is somewhat similar. Here again there are 2 *K*-electrons and 8 outer electrons; in this case the latter may be regarded as being shared in four pairs, according to the theory of G. N. Lewis, between the carbon and the four hydrogen nuclei, the latter being arranged at the four corners of a regular tetrahedron around the carbon nucleus. This structure possesses symmetry of a very high order; the "center of gravity" of the positive charges, like that of the negative electrons, is located at the carbon nucleus, so the forces of electrical interaction between methane molecules are exceedingly small. As in neon, only the weak van der Waals' forces are of importance, and methane is very volatile, with a low heat of vaporization.

The structure of the water molecule presents some entirely different features. Here again there are 8 outer electrons, but only 4 are involved in electron pair formation with the 2 hydrogen atoms. The shape of the water molecule is that of an isosceles triangle; the internuclear O—H distance is very nearly 0.99 Å, and the H—O—H angle is not far from 105°.² The powerful attraction of the oxygen nucleus tends to draw electrons away from the protons, thus leaving the region around them with a net positive charge. The two pairs of unshared electrons tend to concentrate in directions pointing away from the O—H bonds. If a tetrahedron is described about the oxygen atom, with the H nuclei at two of the corners, then regions of high electron density distribution, due to the unshared electrons (centers of negative charge) are concentrated in the directions defined by the other two corners of the tetrahedron. The water molecule is, therefore, an electrically polar structure. In a group of water molecules clustered together, a positively charged region in one molecule tends to orient itself toward a negatively charged region in one of its neighbors. Thus each molecule tends to have four nearest neighbors. Each of its two regions with a concentration of negative charge attracts a proton of a neighbor molecule. Each of its own protons attracts the oxygen end of a neighbor (Fig. 3). Thus each oxygen atom is the center

² One angstrom unit (Å) = 10^{-8} cm. The evidence that the water molecule is triangular and not linear comes from detailed studies of the molecular spectrum, and also from dielectric constant measurements, the significance of which is further considered later in this chapter, and in much further detail in Chapter 6. The same type of evidence shows the carbon dioxide molecule to be linear, $O=C=O$ (see Chapter 10).

of a tetrahedron of other oxygens, the O—O distance (determined from X-ray diffraction studies) being 2.76 Å. The spatial arrangement of an ordered system built up from such tetrahedral groupings is shown more clearly in Fig. 4. This ordered and coordinated structure represents the molecular arrangement in ice, as X-ray studies have shown. The exact position of the protons is not fixed by the X-ray studies, but from the nature of the electrical attractions involved they must lie on (or nearly on) the straight line drawn between adjoining oxygens. They do not lie

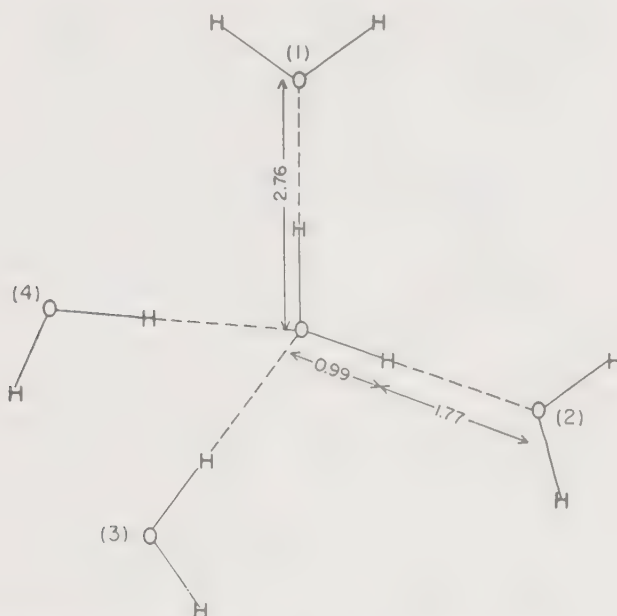


FIG. 3. Tetrahedral coordination of water molecules in ice. Molecules (1) and (2), as well as the central H_2O molecule, lie entirely in the plane of the paper. Molecule (3) lies above this plane, molecule (4) below it, so that oxygens (1), (2), (3), and (4) lie at the corners of a regular tetrahedron. Distances in Angstroms.

midway on these lines, however. As has been said, the O—H distance in a water molecule is 0.99 Å; the O—H \cdots O distance in the ice crystal is 2.76 Å. Thus the distance from a proton to the oxygen in the nearest neighboring molecule is 1.77 Å. This type of linkage, denoted here by a dotted line connecting the proton with a neighboring atom, is known as a hydrogen bond.³ It is very weak compared with the true covalent bond, such as the O—H bonds within a single water molecule. The average energy necessary to break the latter is given by the heat of the reaction



³ Such bonds can be formed only between hydrogen and a strongly electronegative atom, generally F, O, or N. Occasionally hydrogen bonds are formed to carbon (as in liquid HCN) or to sulfur; very rarely to other elements. See Pauling (1940), Chapter IX.

which involves the absorption of 220.3 kcal/mole of water,⁴ or 110 kcal/mole of O—H bonds. The energy of the $\text{H} \cdots \text{O}$ bond has been deduced by Pauling from the heat of sublimation of ice (12.2 kcal/mole), since the process of sublimation involves the breaking of all such bonds, and the effect of other types of intermolecular forces can be estimated from the heat of sublimation of other similar molecules (such as CH_4) which do

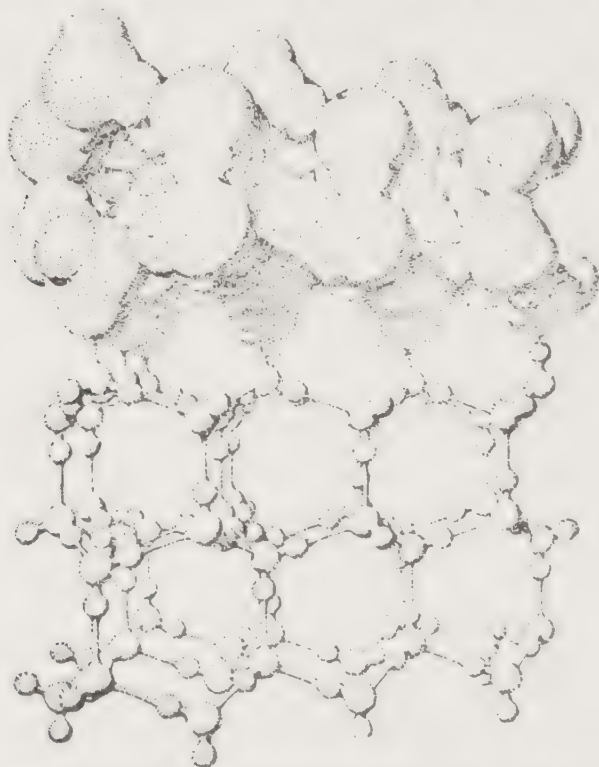


FIG. 4. An extended view of the arrangement of the water molecules in ice. In the lower half of the figure, the hydrogen atoms are shown as small spheres, and the oxygen atoms as slightly larger spheres, to permit the skeleton of the structural pattern to be seen clearly. In the upper half, the actual Van der Waals radii of the molecules are shown, to indicate the packing of the molecules. Note the very open nature of the arrangement. (From L. Pauling, *General Chemistry*, 2nd ed., W. H. Freeman Co., San Francisco, 1954.)

not form hydrogen bonds. Pauling thus estimates the energy of the $\text{H} \cdots \text{O}$ bond in water as approximately 4.5 kcal/mole. This is less than one-twentieth of the value for the O—H bond, but it is sufficient to be a prime factor in determining molecular interaction and orientation in water and other liquids containing hydrogen bonds.

⁴ The kilocalorie (kcal) is equal to 1000 calories. It is the unit of energy commonly employed in metabolic studies, and has frequently been denoted by physiologists as the large calorie or Calorie (with a capital C) to distinguish it from the true calorie. This usage is sometimes confusing. We shall employ the unambiguous symbol kcal to denote it here.

When ice melts, the highly coordinated crystalline structure breaks in many places. This does not mean, however, that all the bonds shown in Fig. 3 are broken. The heat of fusion (1.44 kcal/mole) indicates that only about 15% of them are broken. Cold water contains interconnected groups of water molecules, with structures based on the pattern indicated in Figs. 3 and 4, except that some of the bonds are broken here and there. Such aggregates are constantly forming and breaking up, bonds being broken at one point and new bonds forming at another, so that there are no fixed permanent crystalline structures. As the temperature rises, more bonds are broken, but a considerable proportion of this quasi-crystalline structure remains even at 100°, as shown by the high heat of vaporization of water at this temperature.

The distribution of water molecules in liquid water was studied by Morgan and Warren (1938) by observing the intensity of scattering of X-rays from the liquid as a function of the angle of the scattered radiation relative to the incident X-ray beam. Such angular distribution curves were determined at various temperatures from 1.5° to 83°. Since the X-ray scattering is due essentially to the electrons in the system, rather than to the nuclei, it is possible to calculate the electron density, ρ , in the system as a function of the distance from any of the oxygen atoms, taken as a center. This electron density will be highest immediately around the oxygen nuclei and falls off at greater distances. In crystalline ice, there is a sharp peak of electron density at 2.76 Å, due to the four nearest neighbors of the oxygen atom taken as the center. Since each of these nearest neighbors is surrounded by four others (including the central atom) there will be 12 oxygens at a distance of $2.76(\frac{8}{3})^{1/2} = 4.51$ Å, and other peaks at greater distances. In liquid water these peaks, except the first, are blurred. The first peak, corresponding to the nearest neighbors, has a maximum at 2.90 Å at 15°, increasing to about 3.05 Å at 83°. The peak is also broader and fuzzier than in a true crystal; thus the nearest neighbors are on the average moving further apart as the temperature rises (Fig. 5). The peak near 4.51 is shifted to higher values than in ice; as the temperature rises, the peak fades out and is replaced by a nearly smooth curve, indicating the increasing randomness of the distribution at the higher temperatures. The average number of nearest neighbors is actually greater in water than in ice, although the average distance from the central molecule to its nearest neighbors is increasing. The whole picture fits the view that liquid water has a "broken-down ice" structure, with a constant breaking and re-forming of bonds between neighboring molecules, according to the picture suggested in the preceding paragraph.

We have pictured each proton, in ice or liquid water, as bound to a particular oxygen nucleus by a strong covalent bond, and in general as also attracted to another oxygen nucleus by a much weaker hydrogen

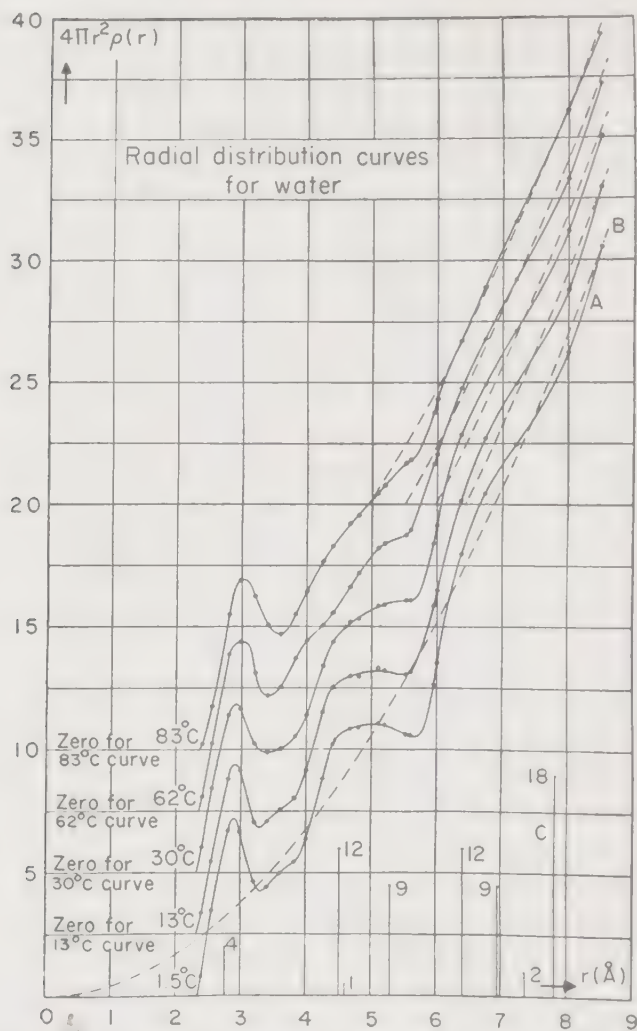
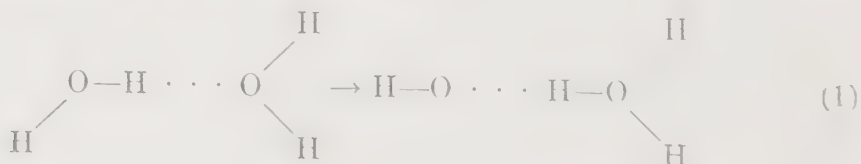


FIG. 5. Electron density distribution in liquid water at various temperatures. Abscissa, distance r from center of an arbitrary oxygen atom in a water molecule. Ordinate, integrated electron density ($4\pi r^2 \rho(r)$) as a function of r . Peaks in the curve correspond to distances at which neighboring atoms tend to congregate. Vertical lines at the bottom give the number and position of neighboring atoms in the ice structure. (From Morgan and Warren, 1938.)

bond. The proton, however, is a relatively light and mobile structure—far more mobile than any other atomic nucleus. There is an excellent chance that from time to time—at intervals indeed of only a very small fraction of a second—it may “hop” from one nucleus to its neighbor. We may represent the reaction by the symbols



This reaction produces two ions, OH^- and H_3O^+ , from two water molecules which are electrically neutral, although polar. We may expect the mean life of such ions to be short. The OH^- ion so formed will probably pick up, almost at once, a proton from a neighbor molecule; the H_3O^+ ion will soon donate one proton to a neighbor. It is possible, of course, that the proton which has made the leap shown in reaction 1 may leap back to its original place. It is more likely, however, that the proton transfers immediately following will involve other protons than that which jumped at first. In any case the electrical attractions involved tend to prevent the formation of large numbers of ions. Conductivity measurements indicate that a kilogram of water at 25° , which contains 55.5 moles of H_2O , contains only 10^{-7} mole of H_3O^+ ions, and of course an equal number of OH^- ions. We may note that the H_3O^+ ion is a typical acid, in the sense that it is a proton donor—it is indeed what is ordinarily called “the hydrogen ion” in aqueous systems. Likewise the OH^- ion is a typical base, that is, a proton acceptor (Chapter 8).

The process of electrical conduction in aqueous solution, for H_3O^+ and OH^- ions, thus proceeds largely by a different mechanism from that involved in conduction of a current by other ions. An ion such as Na^+ or Cl^- , for example, may be regarded physically as resembling a small charged sphere, surrounded by a cluster of oriented water dipoles, more or less tightly bound to it by electrostatic attraction. Under the influence of an applied electric field the ion moves through the medium, among the surrounding water molecules; if the immediately surrounding group of water dipoles is quite tightly bound, it moves also with the ion. The H_3O^+ and OH^- ions may move also in this way. They can in effect, however, move also by a leaping of protons from one oxygen atom to another, by the same process that takes place even in the absence of the field. In the presence of the field, however, such motion does not take place randomly in all directions; the leaps take place more frequently in the direction imposed by the electric field. Since only the light protons actually are transferred, the heavy oxygen atoms remaining at rest, the flow of current by this mechanism can proceed more rapidly than by the other. Thus the electrical mobility of H_3O^+ and OH^- ions is much higher than that of other ions, as is shown by the numerical data recorded later in Chapter 7.

There are two structural features of the water molecule which cause liquid water to be a highly coordinated type of system with strong intermolecular forces: (1) the water molecule is electrically a highly polar structure, this polarity manifesting itself through the formation of hydrogen bonds; (2) the number of protons which form the positive ends of the hydrogen bonds around any given oxygen atom is equal to the number of unshared electron pairs which form the negative ends. This, given the

tetrahedral arrangement of the bonds, determines the formation of extensive three-dimensional structures, with a high degree of cohesion. Molecules such as ammonia and hydrogen fluoride share with water the first of these structural features, but not the second. Thus NH_3 contains three protons, arranged at three corners of a tetrahedron, and *one* unshared electron pair with its electrical charge concentrated in the direction of the fourth corner. The $\text{N}-\text{H} \cdot \cdot \cdot \text{N}$ hydrogen bonds in ammonia are somewhat weaker than the $\text{O}-\text{H} \cdot \cdot \cdot \text{O}$ bonds in water. What is more important, however, is that this structure does not permit the building up of extended linked structures in space. Ammonia molecules may be held together by hydrogen bonds to form chains or rings, but not the type of three-dimensional pattern found in ice or, in a more imperfect form, in liquid water. HF molecules, with one shared electron pair and three unshared ones, suffer from the same sort of geometrical limitation when they are linked together. The $\text{F}-\text{H} \cdot \cdot \cdot \text{F}$ bonds in hydrogen fluoride are even stronger than the $\text{O}-\text{H} \cdot \cdot \cdot \text{O}$ bonds in water—as shown for instance by the fact that HF , even in the vapor state, is largely polymerized; but here again the molecules can associate only in chains or rings. Hence the melting point, boiling point, and heat of vaporization are lower for HF than for water. Hydrogen sulfide, on the other hand, which is so closely analogous in structure to water, has its protons rather deeply buried in the cloud of negative electricity contributed by the electrons from the sulfur atom. Thus the tendency to form $\text{S}-\text{H} \cdot \cdot \cdot \text{S}$ bonds is very weak and influences the properties of the molecule but little. This is indicated by the low melting and boiling points and heat of vaporization of H_2S as compared with H_2O (see Figs. 1 and 2). Also the crystal structure of solid H_2S is entirely different from that of ice; H_2S molecules pack together like spheres, each with twelve nearest neighbors.⁵ This is also the type of structure that is found in solid CH_4 . It indicates the presence only of a general—and in these cases rather weak—intermolecular attraction, without the tendency to form bonds oriented in specific directions.

Thus the structure of water is truly unique. The unique status of

⁵ The type of arrangement involved may be readily seen by examining a pile of shot or of tennis balls. If the balls are arranged in parallel layers, each ball is immediately surrounded by six others in its own layer, three in the layer above, and three in the layer below. According to the manner of juxtaposition of the different layers, two different arrangements, known as cubic close packing and hexagonal close packing, are obtained. In either type of close packing, however, each sphere has twelve nearest neighbors, which is the point that concerns us here. For a further discussion of these structures, see L. Pauling, "The Nature of the Chemical Bond," Chapter X, Section 47; and W. L. Bragg, "The Crystalline State," Chapter VII. H. G. Bell London, 1931.

water, indeed, was clearly recognized more than a hundred years ago in the Bridgewater treatises of Paley and others. It was emphasized early in the twentieth century by L. J. Henderson, on the basis of a much deeper knowledge of scientific facts, and a more critical judgment of the general principles involved. Even Henderson, writing in 1912, preceded the great development of structural chemistry which had its origin in the discovery of X-ray diffraction in crystals, and its interpretation by von Laue, which also began in 1912. The great development of quantum mechanics and of valence theory, in terms of which the properties of molecules can be so largely correlated with their structure, was still to come. At this time, therefore, Henderson was necessarily unaware of the structural considerations that we have described. His analysis was based on the physical properties of water, in comparison with those of other substances, and his discussion of the biological significance of the properties of water remains probably the most important that has been written.

We shall first consider the thermal properties of water. The metabolism of a living organism, which provides the energy for its survival and continued activity, invariably involves the production of heat, and, therefore, raises the temperature. Thus an average man, weighing perhaps 60 kg, may produce in the course of an average day some 2500 kcal of heat. This heat, in a closed system weighing 60 kg, and with the heat capacity of water, would raise the temperature by more than 40°. In the same weight of most other liquids, as we shall see presently, it would produce a much greater rise than this. Yet of course, as we all know, the temperatures of man and other mammals remain nearly constant under ordinary circumstances. Indeed, no animal or plant can tolerate internal temperatures much higher than 37° to 40° for any length of time without lasting damage or death. This appears to be inherent in the nature of life as we know it; proteins, which are essential elements of structure and function in the living organism, possess a very delicately organized type of structure, and many of them are rapidly altered—or, as it is commonly said, denatured—at temperatures not far above 40°. Moreover, the rates of most chemical reactions are very sensitive to temperature, being commonly doubled or tripled by a rise of 10°. The problem of temperature regulation in biology is, therefore, of major importance.

HEAT CAPACITY OF WATER AND OTHER SUBSTANCES

The high heat capacity of water is of primary importance in this regulation. The specific heat of water—that is, its heat capacity per gram—is, by definition, unity. The standard calorie is the amount of heat required to raise the temperature of 1 gram of water from 15° to 16°.

The specific heat of water is very nearly the same at any other temperature from 0° to 100°. Only one other liquid—namely, liquid ammonia, with a specific heat of 1.23—surpasses water in this respect. The specific heat of chloroform is 0.24; of alcohol, 0.6; of hexane and most hydrocarbons, near 0.5. These values are typical of organic liquids in general, very few of which have specific heats higher than 0.7. The specific heats of most solids are even less—that of ice is 0.5; salt, 0.21; glass, 0.20; sugar, 0.30; urea, 0.33; iron, 0.10; and lead, 0.03.

The larger the specific heat, the smaller is the temperature rise produced by a given amount of heat. Consequently, the problem of temperature regulation is far simpler in systems composed chiefly of water or liquid ammonia than in any others. If we wished to construct a liquid thermostat, these would be the most efficient liquids to choose.⁶ These thermostatic properties of water are important not only in the living organism itself, but also in its environment. Especially is this true in the ocean, which is almost certainly the original cradle of living things.

Why do different substances differ in heat capacity? Heat capacity measures the amount of internal energy taken up by the molecules of a substance as the temperature rises. Some of this energy is kinetic energy of translation or rotation of the molecules; some is internal energy of molecular vibration; some is the energy absorption involved in breaking intermolecular bonds. It is in this last type of energy absorption—that due to the breaking of hydrogen bonds—that water and ammonia are particularly rich, and this is largely responsible for their unusually high heat capacities.

HEAT OF VAPORIZATION

Vaporization and condensation are processes of the greatest importance for the maintenance of steady temperatures in systems which are not closed. If a kilogram of water absorbs 1 kcal of heat, its temperature rises by 1°. The evaporation of only 2 grams of the water, however, more than suffices to lower the temperature of the remaining 998 grams to its original level, since the evaporation of each gram (at 40°) involves the absorption of 574 calories. This process is the primary regulating mechanism which controls the temperature of animals and plants and neutralizes the effect of the heat released by metabolic activity. Its effectiveness obviously depends on the magnitude of the heat of vaporization ΔH_v , since material is sacrificed by the organism during the evaporation process, and this material must be replenished if the organism is to survive. The greater the value of ΔH_v per gram, the smaller the turnover of material required.

It is immediately apparent from Fig. 2 that water is vastly superior to

⁶ Ammonia, of course, at 1 atmosphere pressure, is liquid only below -33°.

any of the analogous inorganic compounds in this respect. The data in that figure are plotted as heat of evaporation per mole. If instead we consider the heat per gram, the relative values for H_2O , HF , NH_3 , and CH_4 are not much affected, but the relative superiority of water to all the substances of larger molecular weight—such as H_2S and H_2Se —is even more striking. Water is also superior in this respect to all known organic liquids. A comparison of the data for a few typical liquids is given in Table II.

TABLE II
HEAT OF VAPORIZATION FOR WATER AND CERTAIN OTHER LIQUIDS

Substance	Temperature (°C)	ΔH_v (cal/g)	Substance	Temperature (°C)	ΔH_v (cal/g)
Water	0°	595.9	Acetonitrile	80°	174
Water	20°	584.9	Chloroform	61°	59
Water	40°	574.0	Methyl chloride	20°	95
Water	100°	540.0	Ethyl ether	35°	84
Methyl alcohol	0°	289.2	Ethylamine	15°	15
Methyl alcohol	64.7°	262.8	Acetone	56°	125
Ethyl alcohol	78.3°	204	Hexane	68°	79
<i>n</i> -Propyl alcohol	97.2°	164	Benzene	80°	94
Formic acid	101°	120	Aniline	—	93
Acetic acid	118°	97	Pyridine	115°	101
Propionic acid	139°	99	Piperidine	106°	89
Acetaldehyde	21°	136	Nitromethane	100°	135
Methyl acetate	0°	114	Nitrobenzene	151°	79

ΔH_v is more than twice as large for water as for any other substance on this list; it is more than four times as great for all except the alcohols and acetonitrile. Methyl alcohol—which, except for ammonia, resembles water more than does any other substance—is second to water, but a rather poor second.

This property of water, like its high heat capacity, stabilizes not only the living organism but its environment. The vast energy absorption from the sun by the oceans, lakes, and other bodies of water is—at least in warm climates—for the most part converted into the latent heat of evaporation of water, thereby reducing temperature rise to the minimum. Subsequently, in other and cooler regions of the atmosphere, this latent heat is released by the condensation of water vapor, thereby warming these cooler regions. If we attempt to imagine a dried-up world, it is certain the variations of temperature between night and day would be violent to an extent which is difficult to conceive.

LATENT HEAT OF FUSION

The freezing of water inside a living organism is certainly a rare phenomenon, although certain bacteria and spores have been chilled to

the temperature of liquid air and have been found to be still capable of growth and reproduction when returned to room temperature. The heat released by water on freezing, however, is a significant factor in minimizing the fall of temperature which occurs as winter comes on, in the cooler regions of the earth, in and around large bodies of water. Correspondingly, the heat absorbed by the melting of ice tends to diminish the rate of temperature rise in spring. It is, therefore, significant that the heat of fusion of ice is larger than that of any other substance except ammonia and a few salts (such as KF, NaF, and NaCl) which are held by very strong ionic bonds (Table III).

TABLE III
LATENT HEATS OF FUSION OF WATER, AMMONIA, AND CERTAIN SALTS AND CERTAIN ORGANIC SUBSTANCES

Substance	Melting point (°C)	Heat of fusion (cal/g)	Substance	Melting point (°C)	Heat of fusion (cal/g)
Water	0	79.7	NaCl	804	124
Ammonia	-75	108.1	H ₂ SO ₄	10.3	24
H ₂ O ₂	-1.7	74.1	Acetone	-95.5	23
HNO ₃	-47	9.6	Benzene	5.4	30
LiNO ₃	250	88.5	Acetic acid	16.6	45
CaCl ₂	774	54.3	Methyl alcohol	-97	16
KF	860	108.0	Nitrobenzene	5.7	22
NaF	992	186	Tristearin	56	45

Very few organic substances have heats of fusion greater than 45 cal/g; almost none are higher than 55.⁷ The position of water in this respect, therefore, although not unique, is exceptional. The reason for the high heats of fusion and vaporization of ice and water is clearly to be sought in the system of hydrogen bonds which we have already discussed in detail.

THE EXPANSION OF WATER IN FREEZING

As everyone knows, ice is less dense than water; furthermore the density of water passes through a maximum at 4°. The behavior of water in this respect is not quite unique; a few other substances also expand on

⁷ See, for instance, the tables of heats of fusion given in the "Handbook of Chemistry and Physics" (Chemical Rubber Publishing Co.) or the far more extensive tables given in Landolt-Börnstein's "Tabellen" or the "International Critical Tables." These publications furnish very extensive sources of data for all the physical properties discussed in this chapter.

solidifying; but the phenomenon is extremely rare. Its profound importance for biology and geology has long been recognized.⁸ If ice were heavier than water, it would sink to the bottom on freezing. As Rumford showed about 150 years ago, it is possible to hold a piece of ice trapped at the bottom of a vessel filled with water, and then to heat, or even boil, the water at the top without melting the ice below.

And so it would be with lakes, streams, and oceans were it not for the anomaly and the buoyancy of ice. The coldest water would continually sink to the bottom and there freeze. The ice, once formed, could not be melted, because the warmer water would stay at the surface. Year after year the ice would increase in winter and persist through the summer, until eventually all or much of the body of water, according to the locality, would be turned to ice. As it is, the temperature of the bottom of a body of fresh water cannot be below the point of maximum density; on cooling further the water rises; and ice forms only on the surface. In this way the liquid water below is effectually protected from further cooling, and the body of water persists. In the spring the first warm weather melts the ice, and at the earliest possible moment all ice vanishes.—L. J. Henderson, "The Fitness of the Environment," Chapter III.

We cannot yet completely explain the mechanism of this remarkable phenomenon in all its details, but the general picture of the structure of ice and water, as shown in Figs. 3 and 4, gives a clue. The packing of H_2O molecules in ice can be described as if the molecules were spheres, each with a radius of 1.38 Å (that is, one-half the O—O distance of 2.76), and each with four nearest neighbors. This is a very loose, open type of structure, with a great deal of empty space between the spheres, as may be seen by examining Fig. 4. True close packing of the spheres would cause each one to have twelve closest neighbors instead of four, and if this could be brought about, the density of ice would be between 1.8 and 1.9, instead of being less than unity.⁹ When some of the bonds holding the rigid crystal structure together are broken on melting, the molecules have a tendency to pack closer together to fill up some of the vacant space. This results in increased density. On the other hand, as the temperature rises, the increasing thermal agitation of the molecules tends to make the liquid expand. The first tendency predominates below 4°, the second at higher temperatures. It is obvious that this explanation is incomplete, for it does not tell us in any detail what happens in water

⁸ Henderson's "Fitness of the Environment," p. 107, quotes an interesting "rhapsody" on this subject by Prout in his *Bridgewater Treatise* (1834).

⁹ At high pressures, as the work of Tammann and Bridgman has shown, ordinary ice can be transformed into several other forms of ice (at least five), all of which are denser than ordinary water. The very loose crystal structure of ordinary ice is important in this respect, since it provides space for further rearrangement and compression of the structure. For the conditions of formation of the various kinds of ice, see Chapter 4, Fig. 1.

when the temperature is raised, or why the maximum density should come exactly at 4°; but it does indicate the lines along which a more fundamental explanation is to be sought.

SURFACE TENSION

Water has the highest surface tension of any known liquid, except for certain metals in the liquid state, and certain fused salts, which are so different in structure that we need not consider them in connection with biochemical problems. It is on the whole more illuminating to consider surface forces in terms of surface energy rather than surface tension. When the surface of a liquid is increased, molecules which were formerly in the interior must be brought to the surface. In doing this, work must be done against the attractive forces which operate between these molecules and their neighbors. These forces are exactly the same as those against which work must be done in vaporizing the liquid. In vaporization, however, the molecules are removed almost completely from the

TABLE IV
SURFACE TENSIONS OF CERTAIN LIQUIDS AGAINST AIR

Substance	Temperature (°C)	Surface tension	Substance	Temperature (°C)	Surface tension
Water	0	75.6	Benzene	20	28.9
Water	20	72.75	Chlorobenzene	20	33.2
Water	40	69.56	Cyclohexane	20	25.3
Water	100	58.9	Ethyl acetate	20	23.9
Acetic acid	20	27.6	Chloroform	20	27.1
Acetic acid	50	24.7	Ethyl alcohol	20	22.3
Acetone	20	23.7	Ethyl ether	20	17.0
Ammonia	—	41.3	n-Hexane	20	18.4

intermolecular field of force, whereas for molecules in a surface layer the intermolecular forces remain strong except in the direction pointing toward the air (vapor) phase. Hence it is not surprising that substances with a high heat of vaporization have in general a high surface energy also. This surface energy—that is, the energy which must be expended to increase the surface area by one unit—is given (in ergs per square centimeter) by the values listed in Table IV.¹⁰ This high surface energy of water leads to very high values of the rise of water in the capillary spaces of soil and plants. Surface tension alone is certainly inadequate to explain the rise of water in tall plants and trees; osmotic and other

¹⁰ The physical dimensions of energy (ml^2t^{-2}) per unit area (l^2) are the same as those of tension; that is, of force (mlt^{-2}) per unit length (l), being mt^{-2} in either case (m = mass, l = length, t = time).

forces play a major role. The high capillary rise of water is certainly a profoundly influential factor in biological phenomena, however.

Moreover, as Willard Gibbs showed long ago, the surface energy of any system tends to diminish, if processes which diminish it can occur. Since most substances¹¹ lower the surface energy of water when they dissolve in it, this means that they tend to concentrate at the interface between water and most other phases. This tendency must always be borne in mind in considering the passage of substances across all membranes, and the formation and structure of the membranes themselves.

DIELECTRIC CONSTANT OF WATER

The dielectric constant of a medium may be defined in several ways, all of which are equivalent. We shall consider its significance in detail in Chapter 6. For present purposes it is most convenient to define it in terms of the law of forces between charged bodies which are so small that their dimensions are negligible in comparison with the distance (r) between them; let the charge on one of them be denoted by e_1 , that on the other by e_2 . Then the force between them—reactive if e_1 and e_2 are of like sign, attractive if they are of opposite sign—is:

$$F = \frac{e_1 e_2}{D r^2} \quad (2)$$

Here the factor D in the denominator is the dielectric constant. Its value depends on the medium in which the charges are immersed. It is defined as equal to unity if the medium is a vacuum and is found to be always greater than unity for other media. For a given medium, D varies with the temperature; at constant pressure it always decreases as the temperature increases. Values for the dielectric constants of some typical media are listed in Table V.

The dielectric constant of water far exceeds that of any other pure liquid, except hydrogen cyanide (and perhaps formamide). The implications of this fact are very important, for water is capable of dissolving a far greater variety of salts, acids, and bases—that is, ionic compounds—than any other medium. The dielectric constant of water (Table V) is about forty times as great as that of benzene. If we look at equation (2), it is apparent that the force between, say, a sodium and a chloride ion in benzene, at a given distance of separation, is forty times as great as in water. Thus in benzene the ions of opposite charge tend to rush together;

¹¹ Many inorganic salts, certain amino acids, and a few other compounds form aqueous solutions which have a higher surface energy than pure water. The concentration of such substances, by Gibbs' theorem is *lower* at the interface than in the body of the solution.

soon more ions collect to form a crystal, which settles out from the solution. Hence the amount of sodium chloride which remains in solution in benzene is exceedingly small. In water, on the other hand, the much weaker forces of attraction permit a large number of ions to remain in solution.

Let us consider, on the molecular level, what may be expected to happen when the ions of a salt dissolve in water. We may picture simple monatomic ions as if they were minute charged spheres. The positive charge on the cation tends to attract water molecules around it, the negative (oxygen) end of the water molecule pointing toward the cation. The

TABLE V
DIELECTRIC CONSTANTS OF SOME IMPORTANT LIQUIDS AT 20°

Substance	Dielectric constant
Hexane	1.87
Octane	1.96
Benzene	2.28
Toluene	2.39
Diethyl ether	4.33
Chloroform	5.05
Acetone	21.40
Ethyl alcohol	24.00
Methyl alcohol	33.00
Water	80.00
Hydrogen cyanide	116.00
Glycine (2.5 <i>M</i>) in water	137.00

electric field around the ion is very intense, so the orienting force is very great; a whole cluster of water molecules becomes oriented around the ion, and the intense electric field causes the ions to pack very closely. Similarly water molecules cluster around the anion, in this case with the positive (hydrogen) end of the molecule pointing toward the ion. It may be seen that these oriented shells of water molecules around the ions produce electric fields of their own, and these fields are so oriented as to oppose the fields arising from the ions themselves. Hence the forces of attraction between the ions are weakened; the ions are, as it were, kept afloat and apart from one another by the water molecules which cluster around them. But this is only, on the molecular level, another way of saying that water has a high dielectric constant. Since the presence of ions is characteristic of all biological fluids, and the ions exert a profound influence on many organic molecules, it may be seen how important these dielectric properties of water are.

In order that a substance may have a high dielectric constant, it must be composed of molecules which are electrically asymmetrical (polar). Such polarity is present in water to a high degree; it is also present in many organic molecules such as alcohols, aldehydes, and ketones, and such substances have a much higher dielectric constant than truly non-polar liquids such as the hydrocarbons. In liquid water, however, the effect of the individual dipoles is greatly intensified through their interconnection by hydrogen bonds to form more extended oriented structures. The only important group of substances which are even more highly polar than water are the amino acids, peptides, and proteins (note the effect of glycine on water in the last entry of Table V), and these are, of course, substances of the first importance in biology. We shall return in Chapter 6 to the reasons for their high polarity and shall there consider in much greater detail the methods for measuring the dielectric constant and its significance for molecular structure.

SOLUBILITY

In the preceding discussion we have seen a reason for the high solubility of many salts in water. A vast number of nonionic compounds also dissolve readily in water. Notable among these are compounds containing hydroxyl, carboxyl, amino, or keto groups; all these groups can become interconnected with water molecules by hydrogen bonds. On the other hand, substances containing large hydrocarbon residues are generally only slightly soluble in water, even though they may contain one or more polar groups. Such substances tend to be concentrated in surface films, with the polar group dipping into the surface of the water, and the non-polar hydrocarbon groups clustered together in the adjoining layer above the water surface. The problem of solubility is so vast, however, that we shall here leave it only with this brief mention, returning to the subject from time to time as our discussion of biological systems becomes more searching and more detailed.

REFERENCES

- Bernal, J. D., and Fowler, R. H. (1933). A theory of water and ionic solution, with particular reference to hydrogen and hydroxyl ions. *J. Chem. Phys.* **1**, 515. A paper of fundamental importance on which much of the discussion given here is based.
- Henderson, L. J. (1913). "The Fitness of the Environment," The Macmillan Company, New York. The chapter on water is of fundamental importance.
- Morgan, J., and Warren, B. E. (1938). X-ray analysis of the structure of water. *J. Chem. Phys.* **6**, 666. This gives a detailed analysis of the structure of liquid water, presenting a picture revised in some respects from that of Bernal and Fowler.
- Pauling, L. (1940). "The Nature of the Chemical Bond," 2nd ed., Cornell University Press, Ithaca, New York. Note especially Chapter IX on the hydrogen bond.

- Robinson, R. A., and Stokes, R. H. (1955). "Electrolyte Solutions," Academic Press, New York, and Butterworth's Scientific Publications, London. Chapter 1 gives a useful and interesting discussion of the structure of water, with particular reference to its properties as a solvent for ionic solutes.
- Wells, A. F. (1950). "Structural Inorganic Chemistry," 2nd ed., Oxford, Clarendon Press. Note Chapter XIV on water and hydrates. A useful discussion, giving references to much recent work.

Chapter 3

Problems of Protein Structure

Although water is the major constituent of most biochemical systems, living cells and tissues are far from being ordinary aqueous solutions. They are highly organized systems in which complicated sequences of chemical events are always in progress. The energy released in these processes is not discharged at random but is released in a coordinated fashion so that it is directed to the performance of definite functions—sometimes to the production of mechanical work, as in the contraction of muscle, sometimes to such functions as the formation and discharge of the secretion of a gland. All these highly coordinated functions cannot be maintained without a structural framework, and the existence of such a framework depends on the presence of systems of macromolecules. We shall use the term “macromolecule” to denote molecules with molecular weights larger than a few thousands, and ranging in size up to tens or even hundreds of millions. Such molecules do far more than merely furnishing a structural framework for dynamic processes; the essential catalytic agents which determine these processes are also macromolecules, and these themselves are constantly being built up and broken down.

In some long fibrous structures such as hair, wool, or tendon, molecules in any ordinary sense of the word may be difficult to distinguish at all—these are very long, more or less continuous structures which may extend over distances which are vast compared with ordinary molecular dimensions. By suitable methods of extraction, however, the major protein constituents of some such tissues may sometimes be brought into solution; the resulting preparations, as, for instance, some types of collagen, generally consist of long fibrillar structures. Other macromolecules are more compact, and not far from globular in general shape; for instance, many well-known proteins, such as hemoglobin, ovalbumin, or serum albumin. Often these may be separated quite readily from the tissue in which they occur, and are found to be highly soluble in water or salt solutions, in suitable *pH* ranges generally not far from neutrality.

All the macromolecules referred to above are proteins, which are universally present in living cells and in certain media, such as blood plasma, which surround and bathe the cells. Of comparable importance are the nucleic acids, which in conjunction with certain proteins are the

essential structural elements of the chromosomes, and by inference of the genes, the determinants of heredity, which are contained within the chromosomes. There is suggestive evidence that nucleic acids play an essential part in controlling protein synthesis. Other important macromolecules are the large carbohydrates—cellulose, which is the essential structural framework of most plant cells, and starch and glycogen, which are the storage reservoirs of carbohydrate, available for us as potential energy sources. Because of the extraordinary significance of proteins, however, this chapter will be devoted primarily to them.

We shall omit here many aspects of the chemistry of proteins; discussion of these will be found in other books, some of which are listed at the end of this chapter. Here we are primarily concerned with the dimensions of these molecules and their geometry, their actual sizes and shapes and configurations in space. Also we are concerned with the nature, number, and distribution of the reactive groups within these macromolecules—groups which can accept or donate protons, so that they are actually or potentially ionic, or groups which take part in hydrogen bond formation, thereby stabilizing certain geometrical configurations. Other groups of particular significance will also be considered, such as the sulfhydryl group of cysteine, which is important because of its weakly acidic properties, its reactivity with many metallic ions, and its significance in oxidation-reduction reactions. Its first oxidation product, the disulfide bond which is found in cystine, is of great importance as a cross-link in determining the configurations of proteins.

The biological specificity of these macromolecules is dependent on two factors—the forces of interaction between these reactive groups and their surroundings, and the geometrical arrangements of these groups within the molecules, which make certain interactions possible and interfere with others. These factors are indeed important for all kinds of chemical reactivity, but the geometrical factor becomes particularly important for macromolecules, such as proteins and nucleic acids. The varied and complex topography of the surface of such molecules often imposes stringent requirements on the exact configuration of other molecules which are to be capable of reaction with them. Such specificity is clearly revealed in the interactions of enzymes with their substrates and of antigens with their specific antibodies.

Proteins; Some General Considerations

The natural proteins are exceedingly diverse in structure and function—on the one hand, they include such relatively inert structural components as keratin of wool and hair, or collagen of tendon, and, on the other hand, vast numbers of substances of high chemical and biological

reactivity, including enzymes, viruses, and many hormones, such as insulin. Every protein molecule is a product of a living organism, and these molecules, although they all appear to have a common underlying pattern, are as diverse and as full of individuality as the organisms and tissues from which they are derived. In actively functioning tissues, protein molecules are being constantly synthesized; at the same time—in most adult animal tissues at least, though perhaps not in bacteria—other molecules of the same kind are being broken down, so that a steady state is maintained.

We shall not attempt here to deal with the complex problems of protein metabolism and biosynthesis, nor with the role which is apparently played in the latter process by nucleic acids. This is a field of the most intense interest to biochemists today, but it lies outside the scope of this book. Notable contributions in this field have been made by Rudolf Schoenheimer, D. Rittenberg, F. Lipmann, H. Borsook, E. F. Gale, J. Monod, S. Spiegelman, P. C. Zamecnik, and a number of other workers. The incorporation of amino acids labeled with isotopic carbon, nitrogen, or hydrogen into blood or tissue proteins and the subsequent rate of disappearance of the isotropically labeled proteins have enabled investigators to trace the transformation and exchange of components among the proteins and amino acids of the mammalian organism and to show that they are in a steady state of continual flux. Approximately half the total nitrogen of the liver protein in a rat, for example, is replaced by nitrogen from other tissues and from the diet in about seven days. The mean half-life of the proteins of blood plasma is somewhat longer than this. The hemoglobin of mammalian red cells is far more stable; once the mature red cell has been launched into the blood stream, it survives for about four months, and the hemoglobin molecules within it generally remain intact throughout this period. Relatively inert structural proteins, such as collagen, are transformed even more slowly than this. Even collagen, however, is slowly broken down and replaced. In muscle the structural constituents, myosin and actin, appear to undergo relatively slow transformations, but the half-life of the soluble enzyme proteins of muscle is much shorter, although longer than that of the proteins of blood plasma.

To obtain a protein preparation for study, the chemist must destroy the intact structure of the living cell and interrupt these transformations. Such systems as the blood or the resting seeds of plants, from which many of the most carefully studied proteins have been derived, are relatively inert in this respect. These molecules have emerged from other more active regions of the organism, in which the state of flux and dynamic interchange prevails. In the course of time they return to that flux

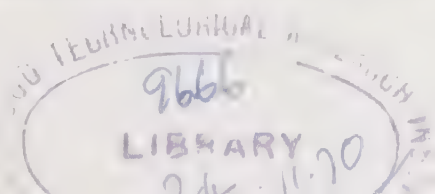
and are engulfed in it once more; but for a limited period they are more stable and can be isolated by suitable techniques with a relatively good chance that the proteins isolated will bear a close resemblance to the molecule in the tissue from which it was derived.

In recent years, protein chemists have been increasingly concerned with the components derived from the actively functioning tissues themselves. Here the problems are more difficult, and the isolated product may be radically altered; and the possibility must always be considered that the product was not present as such in the living tissue before isolation was attempted. It may have formed part of a larger complex in its original state, or it may have reacted to form a new complex when separated from its original associations in the cell. However, the separation from many tissues of crystallized enzymes and other protein constituents, in preparations which show a high degree of biological activity, is one of the great triumphs of science in our time. A beginning has been made with the problems of determining and reconstructing the nature of the integrated systems which function in the intact cell. When the study of such integrated and organized systems is attempted, the biochemist must always be prepared for surprising relations which could not have been inferred merely from the study of purified and separated components.

The art of removing proteins with a minimum of damage from the living tissues in which they occur is a rather special one. Many procedures commonly employed by the organic chemist must be renounced. Protein molecules are complex and delicate structures, and the integrity of the structure of the protein, as it exists in its natural state, is dependent on the linkage of certain parts of the molecule by hydrogen bonds or other weak attachments, which are easily broken. The use of high temperatures, of pH values too far from neutrality, of most organic solvents, or of many quite mild reagents, such as urea, which would have little or no effect on most organic compounds, commonly produces serious disruptive effects on these weak linkages, and leads to the alterations in protein molecules which are commonly denoted by the term denaturation. Some proteins, such as pepsin, are more stable at pH 2 than at pH 7, but the great majority are most stable in nearly neutral solution. Aqueous salt solutions of high or low concentrations, including buffer mixtures, such as acetates, phosphates, or citrates, mixtures of water with organic solvents, such as alcohol or acetone, *at low temperatures*—such reagents as these in general give the most satisfactory results. But even these must be used with caution; the use of highly concentrated salt solutions, for example, can produce serious alterations in the structure of some proteins, such as actomyosin of muscle. Likewise, the effect of even moderate con-

centrations of such substances as alcohol can cause the rapid denaturation of a great number of proteins if the temperature is allowed to rise a little too high. Recently the use of certain ions of heavy metals, such as zinc and cadmium, in low concentrations and at a pH near neutrality, has led to some great successes in the fractional separation of closely related proteins. The development of ion exchange resins, which permit the selective removal of the heavy metal ions from the system after fractionation has taken place, has greatly increased the utility of such methods. In these procedures, however, like all the others, the danger of denaturation of protein must be constantly borne in mind. Even a slight excess of the precipitating agent may on occasion produce far-reaching alterations, especially if the temperature is allowed to rise a little too high or the pH is adjusted to a value which is not optimum for stability. For discussions of the general principles involved and some detailed procedures in the fractionation of proteins, the reader may consult the articles by Edsall (1947), Taylor (1953), and Green and Hughes (1955).

The criteria of purity employed in the study of simpler compounds are, for the most part, quite inapplicable to proteins. Elementary analysis is of some value, but that value is extremely limited for molecules so complex. Melting point determinations are impossible in substances which are profoundly altered, even by moderate heat, and decomposed completely long before anything approaching a melting point is reached. Gradually, however, certain systematic criteria have been evolved for establishing at least the relative purity of a protein preparation. No one of these criteria is in itself sufficient to establish purity, but if a protein preparation is inhomogeneous by any of them, then it is certainly impure. A pure protein preparation should migrate in solution with uniform velocity in an electric field at all pH values in the range within which the molecule is stable; it should move with uniform velocity in a uniform centrifugal field under the same conditions; when it is allowed to diffuse freely, the course of diffusion should follow a certain characteristic pattern. The specificity of immunological reactions for individual proteins can also be used as a delicate indicator for detecting impurities otherwise hard to recognize. A pure protein should show a definite characteristic solubility in any given solvent which does not denature it. The amount of protein dissolved, provided that undissolved protein is present in equilibrium with the solution, should be independent of the amount of undissolved material. The solubility criterion is one of the most delicate of all for determining the purity of proteins; it is probably the closest analog of the melting point test as applied to simpler organic compounds. We should not, however, place too much confidence even in this test;



several per cent of an impurity might well be present in a preparation without giving rise to noticeable deviations from the solubility behavior of a pure chemical individual, as determined by present techniques.

Lately it has been found possible to purify certain relatively small proteins—such as ribonuclease (molecular weight 13,700) and chymotrypsinogen (molecular weight near 20,000)—by passing a solution slowly over a column of a suitable ion exchange resin or other similar material. If the preparation is pure to begin with, the protein emerges in a certain portion of the outflowing liquid. The concentration of protein in the effluent liquid, plotted as a function of the volume of liquid which has flowed through the column, rises to a maximum in a certain region of the curve, and then falls again symmetrically to zero as more liquid flows through. If two or more closely related proteins are present, this symmetry is lost, and the curve may show two or more peaks. Examples of both pure and impure protein preparations, as judged by this criterion, are shown in the experiments represented in Figs. 1*a* and 1*b*. By choosing conditions correctly, it may be possible to separate the peaks completely, and obtain two pure components from the mixture. Thus this technique may be used both to separate certain mixtures of proteins into pure components and to serve as a very sensitive criterion of purity, when purity has been achieved. As yet, it has been very difficult to apply such methods to larger proteins, because of the ease with which they become denatured.

We have mentioned here various criteria of purity without attempting to explain them or to illustrate them in detail. Detailed discussions will be found later in this book, and indeed large sections of some of our later chapters will be devoted to some of the methods to which we have briefly alluded here.

Proteins are so complex, they contain so many reactive groups, that they exist in nature largely in close association with compounds of other types. In practically every cell and tissue, complexes of protein with lipides may be found. Some of these are held together by relatively weak forces and can easily be broken up by extraction with fat solvents; others are much more difficult to break up. Many of the most familiar proteins, such as egg albumin, contain a carbohydrate residue as an integral part of the molecule. One of the most familiar and certainly one of the most remarkable of the proteins, hemoglobin—discussed in more detail in Volume II—depends for its distinctive character on the iron-containing porphyrin group, which is linked to the large globin molecule by characteristic bonds, involving the nitrogenous radicals of one of the amino acids, which may be the imidazole groups of histidine. Here again, the configuration of the amino acids in the globin portion of the hemoglobin molecule is as essential for the distinctive function of the whole

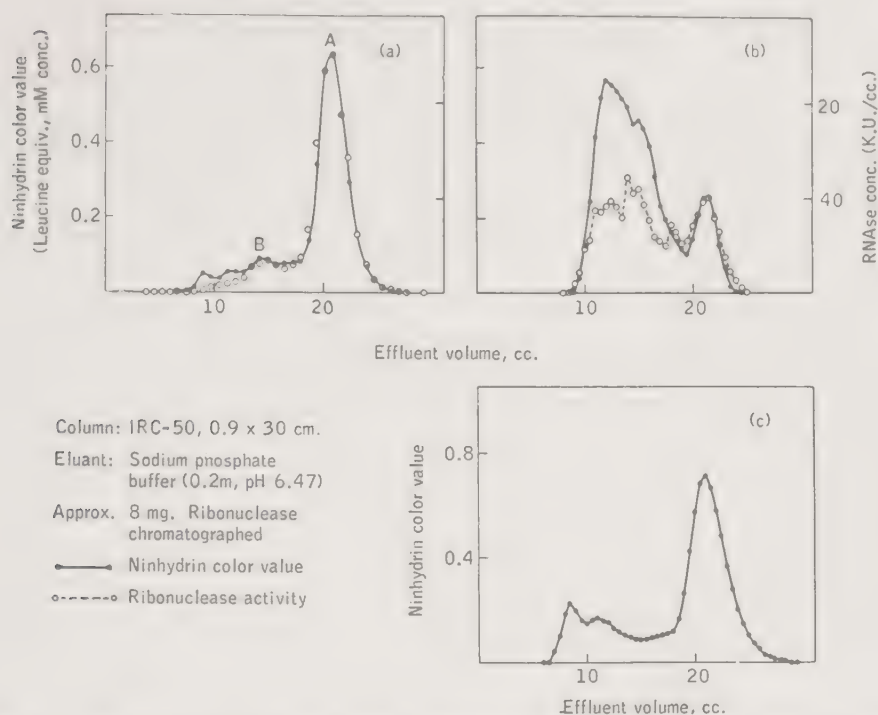


FIG. 1a. Chromatography of several recrystallized preparations of ribonuclease on a column (0.9×30 cm) of the ion exchange resin IRC-50.

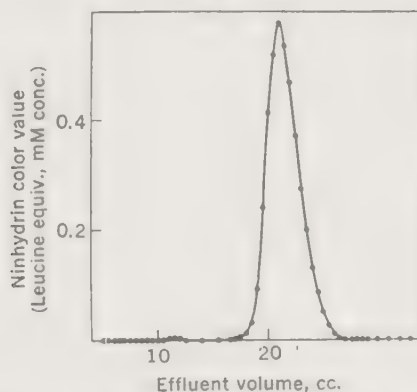


FIG. 1b. Chromatographic homogeneity of purified ribonuclease A, isolated by preparative chromatography from total crystalline ribonuclease. (From Hirs, C. H. W., Moore, S., and Stein, W. H., *J. Biol. Chem.* **200**, 493, 1953.)

structure as is the iron-containing group, which combines directly with oxygen. The same group, combined with other systems of amino acid residues, produces very different molecules, each with a characteristic function, such as the cytochromes—nearly universal constituents of the living tissue and essential as oxidation catalysts—and catalase and peroxidase, each with its specific catalytic function. Even such molecules as the carotenoids, for which there seems no obvious basis for a structural

combination with protein, can exist in close association with protein molecules. This is perhaps most clearly exemplified by the work of Wald and his associates, which has demonstrated the intimate association of vitamin A with the protein visual purple in the retina (see, for instance, Wald, 1953, 1954). The nucleoproteins in which proteins and nucleic acids are closely intertwined are probably the most important of all the conjugated proteins.

For some purposes, the aim of the chemist is to isolate these structures as nearly intact as possible; for others, to break them down into their components. The ultimate aim—the reconstruction of the total system in terms of constituents of known structure—remains the same, but that goal is still rather distant.

At this stage, however, it would take us too far afield to discuss the conjugated proteins. The problem of the structure characteristic of all proteins, the linkages of amino acid residues, is quite complicated enough in itself to occupy us here. Many proteins with definite and specific functions, such as the hormone insulin or the proteolytic enzymes of the digestive tract, are themselves proteins, composed entirely of amino acid residues so far as we know.

Amino Acids and Peptides as Dipolar Ions

All the amino acids derived from protein hydrolysis are α -amino acids with the general structure $^+\text{H}_3\text{N}\cdot\text{CH}(\text{R})\cdot\text{COO}^-$ in neutral solution. (Proline and hydroxyproline, which are closely related structures, are discussed below.) We write this structure so as to indicate the presence of electrically charged groups rather than employing the conventional formula of the isomeric, uncharged molecule, $\text{H}_2\text{N}\cdot\text{CH}(\text{R})\cdot\text{COOH}$, since we wish to emphasize the ionic character of the groups in these molecules at neutral $p\text{H}$. The indicated structure contains simultaneously a cationic ammonium group and an anionic carboxyl ion; it is a dipole of very high electric moment, about eight times as great as that of a water molecule. Peptides such as glycylglycine ($^+\text{H}_3\text{N}\cdot\text{CH}_2\cdot\text{CONH}\cdot\text{CH}_2\text{COO}^-$) and amino acids containing more than one carbon atom separating the ammonium and the carboxylate groups give still higher electric moments. Such a structure, containing ionic groups of opposite charge in equal numbers, is known in German as a *Zwitterion* and in English is commonly called a dipolar ion. The interactions of dipolar ions with other substances are considered in Chapter 5, and the evidence concerning their dipole moments, and the dielectric constants of their solutions, in Chapter 6. Actually dipolar ions and the isomeric uncharged molecules are in equilibrium in aqueous solutions; the nature of this equilibrium depends on the acid and basic strength of the ionizing groups involved and is considered in detail in Chapter 9.

In general, throughout the discussion in this book, we shall emphasize, in writing chemical formulas, the state of ionization of ionizable groups. Formulas written in this fashion provide more information than formulas as commonly written. For instance, it is known that dipolar ions crystallize in such a way as to bring a positively charged group of one molecule into close proximity to a negatively charged group of one of its neighbors, the two being generally linked by a hydrogen bond, involving one of the hydrogens of an —NH_3^+ group, to a —COO^- group of the neighboring molecule. This type of linkage, involving a strong electrostatic attraction superimposed upon a hydrogen bond, is much stronger than the ordinary bonds between organic molecules in crystals. Much energy is required to disrupt them. Correspondingly, the melting points of amino acids, peptides, and other dipolar ions are very high—commonly of the order of 300° . Indeed these substances commonly decompose when, or before, they melt, indicating that the intramolecular covalent bonds, strong as they are, are broken at high temperatures about as readily as the intermolecular bonds, which are reinforced by strong electrostatic attractions.

Also, the presence of ionic groups, as indicated in the general structure shown above, immediately suggests that such a molecule must have a strong affinity for water, because of the intense electrostatic forces around the ionic groups which attract surrounding water molecules. Hence, because of their high dipole moments and the capacity of both the —NH_3^+ and —COO^- groups to attach themselves through hydrogen bonds to water molecules, dipolar ions in general dissolve much more readily in highly polar solvents like water than in slightly polar or nonpolar solvents of low dielectric constant. Aqueous solutions of dipolar ions have very high dielectric constants, often much higher than that of pure water, because of the extremely high electric moments of the dipolar ions (Chapter 6). Dipolar ions are, therefore, much more soluble in water than in organic solvents, unless they contain very large nonpolar side chains, which increase the affinity for organic solvents. Even a molecule with as large a nonpolar side chain as that of phenylalanine is considerably more soluble in water than in ethanol and is almost completely insoluble in ether, benzene, or hexane. Dipolar ions are also distinguished from their uncharged isomers by characteristic differences in the vibrational frequencies of the acidic and basic groups, in the charged or uncharged state. These vibrational frequencies may be detected by means of infrared and Raman spectra (see, for instance, Cohn and Edsall, 1943, Chapter 1).

All these characteristics of dipolar ions—high melting point, high electric moment, high positive dielectric increment in water and other solvents, high solubility in water relative to that in less polar solvents,

characteristic acid and basic properties, and characteristic spectroscopic frequencies—are recalled, to the experienced chemist, by the formula $+H_3N\cdot CH(R)\cdot COO^-$. A chemical formula should convey in brief, explicitly or implicitly, as many as possible of the properties of the compound denoted. The amount of meaning to be derived from a formula of course increases with the breadth of knowledge of the man who reads the formula; but the formula itself should also be chosen so as to convey, in a simple way, as much meaning as possible.

Characteristics of the Amino Acids Derived from Proteins

On hydrolysis under suitable conditions, the proteins are broken down into α -amino acids. The most generally employed method of hydrolysis is with boiling strong acid, such as 6 *N* hydrochloric acid. If carbohydrate is present, tryptophan is largely destroyed by this treatment and there may be some loss of cysteine and cystine also. For determination of these amino acids, hydrolysis with alkali may be useful. Hydrolysis with proteolytic enzymes may also be employed; this type of hydrolysis is, of course, carried out at moderate temperatures (37° or below) and involves a minimum of decomposition of the amino acids themselves. No one enzyme, however, will break a protein down completely into amino acids; even with a carefully chosen combination of several enzymes, specific for different kinds of linkages, some of the bonds between amino acid residues generally remain unbroken. Acid hydrolysis, therefore, remains generally the most suitable for analytical studies of the relative numbers of most of the different amino acids in a protein hydrolyzate.

Historically the separation from one another of all the amino acids in such a hydrolyzate has been a most formidable problem. The amino acids, because of their very strong interactions with one another and with the solvent water, due to their dipolar ionic structure, are quite involatile, and even when one of them can be induced to crystallize it is likely to be contaminated with impurities from the many other constituents of the system. The separation of pure amino acids from protein hydrolyzates was a process that extended over more than a century, before all the now recognized amino acids were separated and identified. The highest skill of many of the greatest chemists was called forth in the achievement of these separations. The history of these events has been well recorded by Vickery and Schmidt (1931); since they wrote, two more amino acids, threonine and hydroxylysine, have been discovered—the former by W. C. Rose and his associates, the latter by D. D. Van Slyke and his associates. The list of natural amino acids derived from the best-known proteins now appears to be complete, since the quantitative analysis of a large number of proteins by modern methods leaves nothing to

be accounted for, within the limits of experimental error, when the content of all the known amino acids is added up. Special and unusual proteins from particular organisms, however, may still be found to contain amino acids as yet unknown.

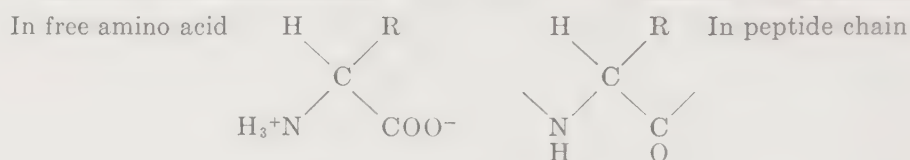
Today, largely owing to the development of chromatography and of modern ion exchange resins, the separation of the components of a complex mixture of amino acids is—at least on a small scale—a very simple and straightforward process. A complete amino acid analysis may now be carried out on the hydrolyzate from a few milligrams of protein, and the whole process may take two or three days. Less than twenty years ago, a similar undertaking would have required perhaps 100 grams of protein, perhaps considerably more. Months of labor would have been involved, and the final result would have been far less accurate than the short and simple procedure referred to above. We return later to a further discussion of the methods involved.

A list of the amino acid side chains—the R groups in the formula $+H_3N\cdot CH(R)\cdot COO^-$ —is given in Table I. The list should already be familiar to most readers of this book, but some comments are appropriate.

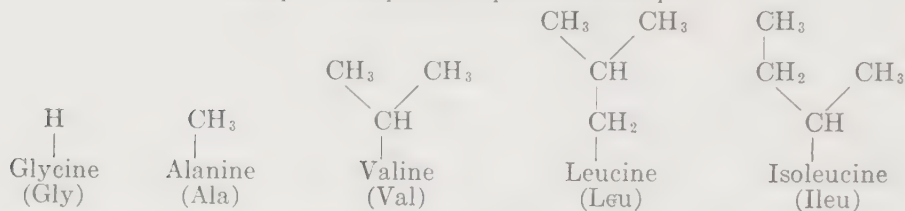
1. All the amino acids on this list are α -amino acids, except for proline and hydroxyproline which are α -imino acids. This is not accidental; the pattern of the protein molecule is built upon organized arrangements of polypeptide chains, which often are coiled about a helical axis by hydrogen bonds within the chain, or else are held together by hydrogen bonded cross-links between more extended peptide chains lying side by side. In either type of structure, there is a recurring pattern that repeats at regular intervals; this regularity would be destroyed, and the stability of the structure gravely impaired, by any change in the repeating distance along the peptide chain, such as would be introduced by the presence of a β - or γ -amino acid residue at intervals here and there. Amino acids other than α are indeed found in nature; β -alanine, for example, is found as a free amino acid, and also combined with other residues in carnosine, anserine, and coenzyme A, but it is never found in proteins as part of a long peptide chain.

2. With the single exception of glycine, all the amino acid residues listed in Table I have a center of asymmetry at the α -carbon atom. Prolonged research has shown that the arrangement of the bonds around the α -carbon—the H atom, the carboxylate group, the ammonium group, and the R group—is the same in every case for all amino acids derived from proteins. This spatial arrangement is known as the L-configuration. Some of the L-amino acids are levorotatory in solution, some dextrorotatory; some are dextrorotatory in one state of ionization, but levorotatory in another. The sign of the optical rotation in itself gives little clue to

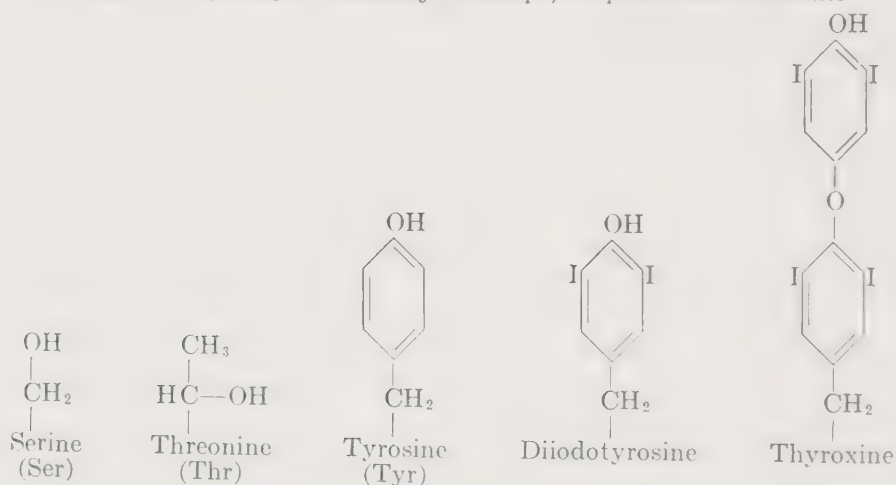
TABLE I
STRUCTURE OF SIDE CHAINS (R) ON AMINO ACIDS FOUND IN PROTEINS



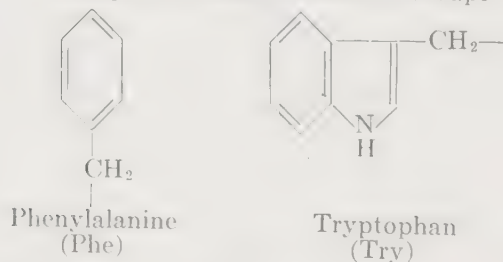
Group I. Nonpolar Aliphatic R Groups



Group II. Hydroxyl-Containing R Groups, Aliphatic and Aromatic



Group III. Other Aromatic R Groups



Group IV. Aspartic and Glutamic Acids and Their Amides

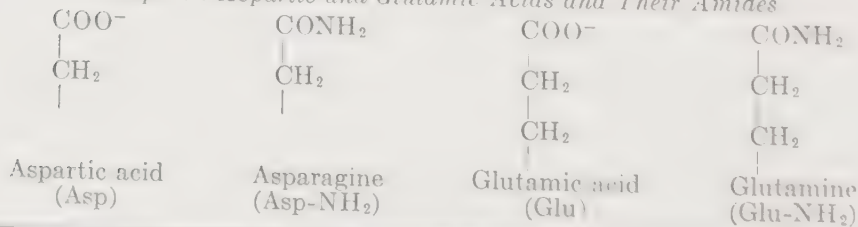
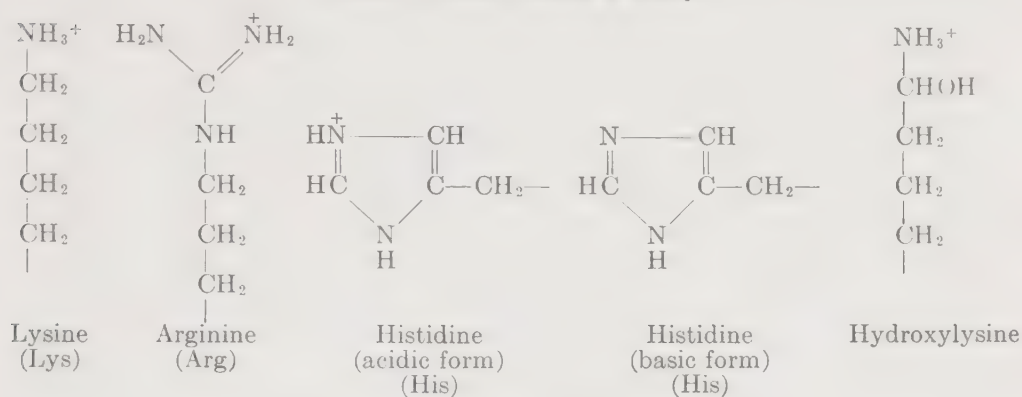
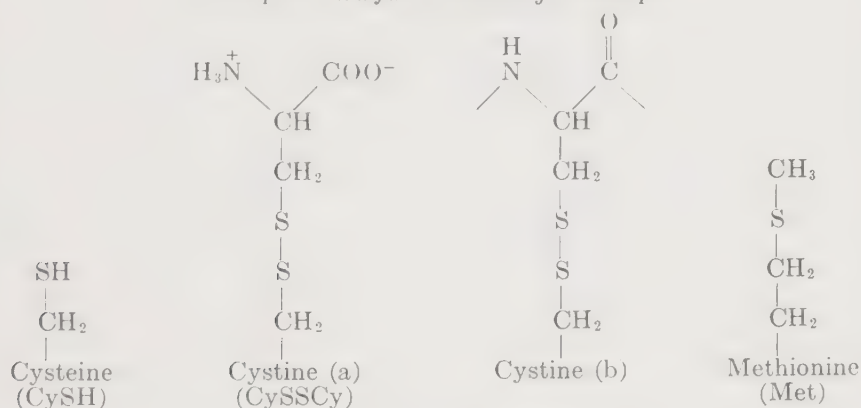


TABLE I (Continued)

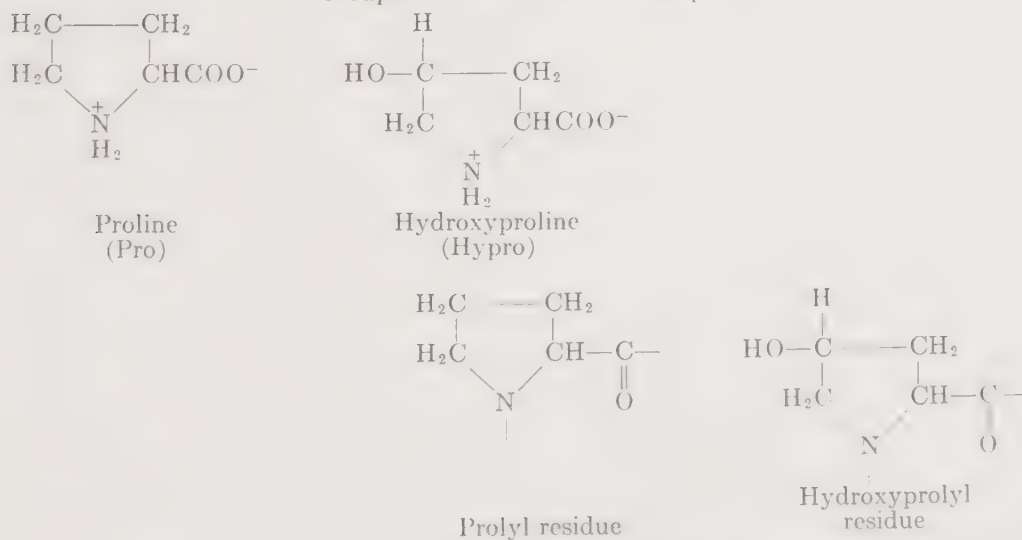
Group V. Basic Amino R Groups



Group VI. Sulfur-Containing R Groups



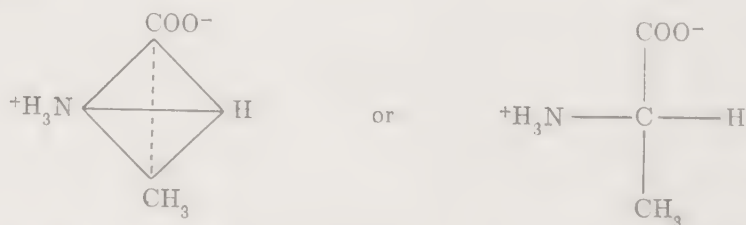
Group VII. Imino Acid R Groups



the actual steric configuration, although it is a fairly general rule that the cationic form of an amino acid—the form present in strongly acid solution—has a more positive, or less negative, value of the optical rotation than the dipolar ion (Clough's rule). Definite evidence for the steric interrelation of the L-amino acids with one another, and with other related compounds such as L-lactic acid, has been established only by a long series of chemical interconversions, carried out in such a way as to assure that inversion of configuration around the α -carbon atom has not occurred at any step of the process (see, for instance, Neuberger, 1948).

Many of the optical enantiomorphs, the amino acids of the D-series, have been found in nature. Thus a C-dimethyl derivative of D-cysteine is present in the penicillins; the antibiotic Gramicidin-S from *B. brevis* contains a residue of D-phenylalanine, along with several L-amino acid residues; and the capsular substance of *B. anthracis* and related organisms contains a large polypeptide made up of D-glutamic acid residues linked together. As yet, however, there is no clear evidence that D-amino acid residues ever exist in a true protein. Small amounts of such residues can indeed often be detected in acid hydrolyzates of proteins; but, by comparison with hydrolyzates from synthetic peptides of known structure and configuration, it is generally concluded that these traces of D-residues arise from racemization occurring during hydrolysis. Alkaline hydrolysis is well known to lead to extensive racemization. Enzymatic hydrolysis has apparently never led to the liberation of any free amino acids from proteins except the pure L-forms.

Recently J. M. Bijvoet and his colleagues in Utrecht have established, by a novel type of X-ray diffraction technique, the absolute spatial configuration of the tartaric acids, and have shown that the usual Fischer convention for representing such configurations is actually correct. Because of the chemical interrelations previously established, by chemical transformations between different classes of compounds, this work also establishes the absolute configuration of the L-amino acids. The configuration of L-alanine, written according to the Fischer convention, is shown below:



To leave no possible ambiguity as to the arrangement denoted here, we may point out that, if one chooses the hydrogen atom attached to the

α -carbon as the apex of the tetrahedron shown in the left-hand diagram above and looks out from this atom onto the other three groups, they will appear in the order methyl, amino, carboxyl, as one proceeds in the clockwise direction. For D-alanine, they will appear in this order on proceeding in the counterclockwise direction.

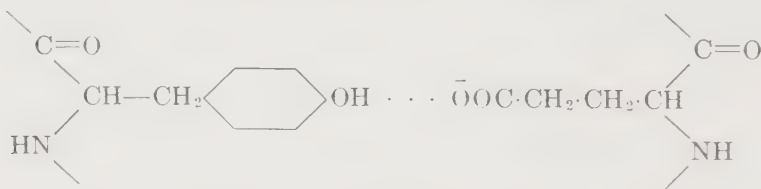
Four of the natural amino acids shown in Table I—threonine, isoleucine, hydroxyproline, and hydroxylysine—contain a second asymmetric carbon atom. For hydroxyproline, Neuberger (1948) has shown that in the natural hydroxy-L-proline the hydroxyl and the carboxyl groups are *trans* with respect to the plane of the pyrrolidine ring. Concerning the configurations of the other three amino acids with two asymmetric centers, the review of Greenstein (1954) may be consulted.

3. The list given in Table I includes twenty-five different R groups (two of these, histidine and cystine, are represented in two different forms). Four—diiodotyrosine, thyroxine, hydroxyproline, and hydroxylysine—have a very restricted distribution, the first two being seldom or never found except in the thyroid gland and the hormones produced by it, and the last two in collagen and its breakdown products. Many books do not list asparagine and glutamine separately, as amino acid residues distinct from aspartic and glutamic acids. In ordinary acid and alkaline hydrolysis the amide linkages in these residues are broken with great rapidity, giving ammonia and free aspartic or glutamic acid in the final hydrolyzate. Careful enzymatic hydrolysis, however, releases asparagine and glutamine in large amounts, showing that these residues were present in the original protein.

4. The formulas for the side chains, as written in Table I, indicate the state of ionization of ionizable groups, as they exist at pH 7. The imidazole group of histidine ionizes near this pH and may, therefore, be either positively charged or uncharged under physiological conditions. The α -amino groups which may be present at the ends of peptide chains may also be either charged or uncharged. The ϵ -ammonium groups of lysine and the guanidinium groups of arginine residues are essentially all in the positively charged state at pH 7, whereas virtually all the carboxyl groups are negatively charged at this pH. These statements hold in so far as such groups are actually free to ionize in the state in which they exist within the protein structure. (For further details, see Chapters 8 and 9.) In some proteins, some of these ionizable groups may be so tied up, by hydrogen bonding with other groups for instance, that they do not ionize freely as such groups do in simpler compounds. For steric reasons also, some groups may be unreactive because access to them is blocked by neighboring groups. Nevertheless, protein molecules, at any pH, contain large numbers of charged groups in the side chains of the amino acid residues, and

the presence of this system of electric charges profoundly influences the properties of the proteins. It is, for instance, probably the major factor in determining their high solubility in water relative to that in nonpolar solvents, as it is in the amino acids and peptides.

The phenolic hydroxyl groups of the tyrosine residues are weakly acidic and practically un-ionized near pH 7. In diiodotyrosine and thyroxine, however, the presence of iodine atoms on the carbons immediately adjoining the hydroxyl greatly increases the acidity of the hydroxyl, so that it may be largely ionized even at neutral pH , as we shall see in Chapter 8. Of more general significance, however, is the fact that the tyrosine hydroxyl groups in many proteins may enter into hydrogen bond formation with suitable acceptor groups in the side chains of other amino acid residues, if the spatial relations are favorable. A frequently suggested type of interaction of this sort is with ionized carboxyl groups to form the link



Hydrogen bonds of this sort may, under suitable conditions, be quite strong, owing to the attraction of the negative charge on the carboxylate ion for the hydroxyl hydrogen, and they may remain unbroken over a very wide pH range, as we shall see in Chapter 9. Different proteins differ greatly in this respect. In ovalbumin of hen's eggs, for instance, the hydroxyl groups of the tyrosine residues appear to be bound in some way—perhaps through hydrogen bonding of the type indicated above—so that they are not free to ionize until a very alkaline pH is attained. Ionization occurs, after the reaction of the solution has become sufficiently alkaline, as shown by a change of ultraviolet light absorption. Once the change has occurred, the bond which inhibited ionization appears to be broken, and the tyrosine hydroxyl groups then behave as they do in simpler compounds. In other proteins, such as insulin, most of the tyrosine hydroxyl groups appear to be free even in the native protein. The aliphatic hydroxyl groups of serine and threonine can also enter into hydrogen bond formation, although they are far more weakly acidic than the phenolic groups, and there is as yet little or no direct evidence that they actually do form such bonds in protein structures.

The sulfhydryl groups of the cysteine side chains are also weakly acidic, slightly more so than the phenolic hydroxyl groups of tyrosine. In simpler sulfhydryl compounds they are readily oxidized by many

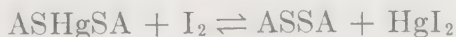
oxidizing agents to form disulfide groups, sulfoxides, sulfones, or sulfonic acids. They also combine readily with the ions of mercuric mercury, copper, silver, and other metals. The sulfhydryl groups of proteins are relatively few in number; many proteins indeed contain none at all, such as insulin which contains three disulfide bonds of cystine in a molecule of molecular weight near 5700, but no free cysteine sulfhydryl group. Serum albumin, a molecule approximately ten times as large, contains only one detectable sulfhydryl group, but about seventeen disulfide linkages. This molecule undergoes a remarkable reaction with mercury, discovered by Hughes (1946). Denoting the albumin, with its one sulfhydryl group, by the abbreviated symbol ASH, we may write the first step in its interaction with a mercury salt, such as mercuric chloride, as



This reaction is to be expected; it is typical indeed of any mercaptan; but it is followed by a second reaction with a second molecule of albumin:



This links the two molecules of albumin together through a mercury atom which connects their two sulfhydryl groups. The resulting mercury dimer of albumin forms excellent crystals, which were obtained by Hughes. Every step in the process is reversible. This stable yet reversible linkage of two large protein molecules by a single chemical bond of this sort is a reaction which has not yet been duplicated in other proteins. Cautious treatment of the crystals of mercury dimer with iodine (Straessle, 1954) oxidizes the sulfhydryl groups to disulfide:



This disulfide dimer of albumin has not yet been obtained in pure form.

The elongated muscle protein myosin, so important in the contractile process, contains a number of reactive sulfhydryl groups. These are of the utmost importance in contraction. It has been found that the contractile process in actomyosin threads, or in simple muscle fibers, is regularly associated with the breakdown of adenosine triphosphate (ATP). Both the splitting of ATP and the contractile process are brought to a halt if the —SH groups of actomyosin are blocked by the addition of an organic mercurial. If the mercurial is removed, for instance by the addition of another sulfhydryl compound such as cysteine, both processes start again. Many other important enzymatic processes are dependent on the presence of reactive —SH groups in the protein structure of the enzyme; the affinity of hemoglobin for oxygen is markedly altered by the attachment of an organic mercurial to the two adjoining sulfhydryl

groups which are apparently located close to the heme group which contains the iron atom that combines directly with the oxygen. (Riggs and Wolbach, 1956.)

Thus, although the sulfhydryl groups of proteins are numerically a very small fraction of the total number of amino acid residues, they often play a role of first-rate importance in the functioning of the protein. Often, it should be added, such groups are not detectable in native proteins by the usual chemical tests for sulfhydryl groups, such as their reactivity with iodine, iodoacetamide, or various mercurials; only after the protein has been treated with a denaturing agent, such as an alkyl sulfate, or a concentrated solution of urea or guanidine hydrochloride, do such —SH groups become reactive. There are great variations in this respect from protein to protein. Thus the —SH groups of ovalbumin of hen's eggs are readily titrated in the presence of denaturing agents—about five such groups per molecule have been found, based on a molecular weight of 43,000—but they fail completely, in the undenatured protein, to react with such reagents as ferricyanide. The enzyme urease from jack bean contains some sulfhydryl groups which react readily in the native protein, and others which react only very slowly, or only in the presence of a denaturing agent. Some, at least, of these groups are essential to the action of the enzyme on urea.

This lack of reactivity of certain groups in native proteins is not a unique characteristic of sulfhydryl groups. The similar lack of reactivity of the tyrosine hydroxyl groups in many proteins has already been mentioned. Likewise, in certain proteins such as β -lactoglobulin, some of the ϵ -amino groups of the lysine residues are unable to react with a reagent such as 2,4-dinitrofluorobenzene to form dinitrophenyl derivatives, although they become quite reactive after the protein has been denatured by the action of heat, acid, alkali, or concentrated urea solution. With regard to all these groups there is great variation from one protein to another. According to the plan of the particular protein molecule, some or nearly all of the side chains of one or another type of residue may be unreactive toward certain reagents in the native protein, whereas in another protein, built on a different scheme to perform a different kind of function, all such groups may be free and reactive.

Some comment is called for on the nonpolar side chains of proteins—notably those of alanine, valine, leucine, isoleucine, and phenylalanine. These are very unreactive toward most reagents which protein molecules are likely to encounter, by comparison with the charged and polar groups. They are likely to serve as a source of attraction for other nonpolar groups in neighboring molecules, since nonpolar residues, immersed in a polar medium, tend to associate with one another rather than with the

polar molecules around them. The significance of the nonpolar residues in proteins must be more specific than this, however; each one must serve a particular function in its own place within the protein structure, although at present we have scarcely the beginning of a conception of what this function is. It is interesting that the residues found in proteins do not include any of the longer straight-chain aliphatic residues—so far as we know, α -amino-*n*-butyric, α -amino-*n*-valeric, and α -amino-*n*-caproic acid residues are never found in proteins, although they have sometimes been claimed on insufficient evidence. Yet the branched-chain isomers of the latter two—valine, leucine, and isoleucine—are among the most common and the most universally present of all the amino acid residues. We may suspect that the particular set of amino acid residues employed in the construction of proteins throughout nature has not been selected at random from all the vast variety of possible R groups that conceivably might have been chosen, but that fundamental structural considerations, of a sort not now clearly apparent, have determined the selection, over the long course of biochemical evolution.

Amino Acid Analysis of Proteins

As a preliminary to more detailed knowledge of the structure of a protein, analysis of the total content of amino acids present is important. The usual preliminary to such an analysis is a hydrolysis of the protein with strong acid—for instance 6 *N* hydrochloric acid at 100° for varying periods of time. This breaks the protein down completely, or nearly completely, into its constituent amino acids; a few of the linkages holding amino acid residues together, such as those involving valyl residues, are split much more slowly than most of the others. They may require 72 to 96 hours for complete splitting under the above conditions, whereas the great majority of the free amino acids are released within 24 hours. Tryptophan is largely destroyed by acid hydrolysis, but is stable during hydrolysis in hot dilute alkali, and must be separately determined. Serine and threonine are also partially destroyed on hydrolysis with acid; the loss is roughly about 10%, and this must be corrected for.

This mixture of amino acids in the acid hydrolyzate must then be resolved into its constituents, and the amount of each constituent determined by a suitable analytical method. Very high resolving power is obtained by the use of chromatography on paper, or with suitable ion exchange resins. Both methods are widely used, and are in a general way familiar to nearly all chemists, but the underlying theory of the operations involved is by no means completely developed as yet.

In paper chromatography, developed in England by A. J. P. Martin and others, a small drop of the solution to be studied is placed on a piece

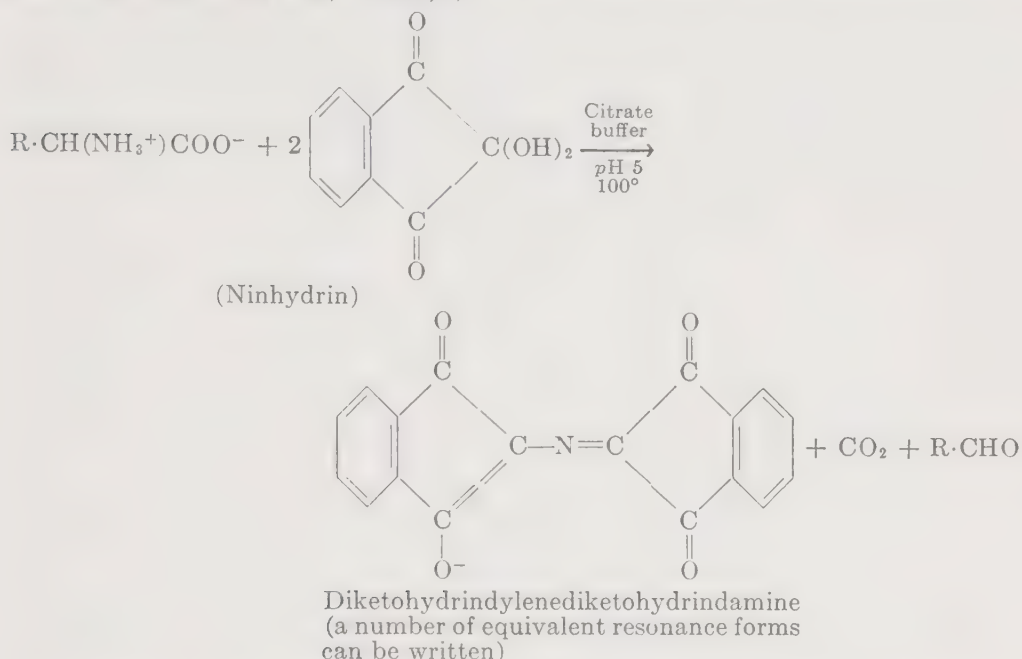
of moist filter paper. The end of the filter paper adjoining the drop dips into a reservoir containing an organic liquid, such as butanol or phenol, which slowly percolates through the paper. The components of the amino acid mixture move in the same direction as the front of the advancing organic solvent, but less rapidly, each with its own characteristic velocity. The ratio of the velocity of a particular constituent to that of the advancing liquid front is generally denoted by the symbol R_f ; it is, of course, always a number less than unity. Substances with nonpolar side chains are more soluble in organic solvents, relative to water, than substances with charged or polar side chains (see Chapter 5); hence they generally have larger R_f values. The process may be visualized as involving a constant exchange of each kind of amino acid molecule between the two liquid phases—the water which is held in the filter paper, and the organic solvent moving through it. By a vast number of repeated exchanges, a small difference in relative solubility in the two media between different amino acids may be magnified into a large difference in rate of flow along the paper; such a process is somewhat equivalent to a fractional distillation with a great many steps. Actually this picture is somewhat oversimplified, for it implies that the paper acts merely as an inert framework to hold one of the two liquid phases, while the other flows relative to it. In fact, however, the data cannot be explained entirely in these terms; there is reason to believe that the amino acids are to varying degrees reversibly adsorbed on the paper, and that this adsorption, as well as the partition ratio between the two liquid phases, affects the relative rates of flow.¹

Often treatment with a single solvent is not sufficient to resolve all the different components in a complex mixture of amino acids or other similar compounds. In such cases, the paper may be dried, again moistened, and then exposed to the action of another solvent, the direction of flow in this second operation being at right angles to that in the first. By suitable choice of the two successive solvents, it is possible to resolve all the amino acids in a protein hydrolyzate into separate spots on the paper. Generally the hydrolyzate from a fraction of a milligram of the original protein suffices to give a complete pattern.

The positions of the spots are commonly detected by spraying the paper with ninhydrin (indane-1,2,3-trione-2-hydrate). The reaction (for

¹ Paper indeed possesses a dual structure (Martin, 1950; Moore and Stein, 1952). About half the weight of paper soaked in water is somewhat like an amorphous gel; the rest is the fibrous network which supports the gel and maintains the structure. Adsorption of solutes may occur on this network, and also to some extent on certain constituents of the gel-like portion of the paper. The latter acts largely, however, as an aqueous phase which exchanges solutes with the surrounding organic phase in much the same way as during equilibration in a separatory funnel.

all α -amino acids except cysteine) yields the same blue-colored product, with an absorption maximum at $570\text{ m}\mu$; carbon dioxide and an aldehyde characteristic of the amino acid are also liberated (see, for instance, Moore and Stein, 1948, 1954a,b):



The different spots can be identified by comparison runs with known amino acid mixtures, with known relative rates of flow. When located, each spot can be cut out, the amino acid extracted with water, and a quantitative (or semiquantitative) determination of the amount of material in each can be carried out by determining the color developed with ninhydrin, or in other ways. Several valuable books and review articles present the subject in detail (Turba, 1954; Lederer and Lederer, 1953; Block *et al.*, 1955; Thompson and Thompson, 1955). A very useful method which gives quite good quantitative results on protein hydrolyzates was developed by Levy (1954). It involves conversion of the amino acids into their N-2,4-dinitrophenyl derivatives, which are yellow and easily identified, before the hydrolyzate is subjected to paper chromatography. These dinitrophenyl (DNP) derivatives are further discussed below, in connection with the study of sequences of amino acid residues in peptide chains.

A method of resolution which is somewhat more elaborate, but capable of greater accuracy, is that of chromatography on ion exchange resin columns (Moore and Stein, 1951, 1954a).² The resins employed were

² Columns of starch were employed earlier by the same investigators, but the exchange resins proved to give better results more simply.

sulfonated polystyrenes (Fig. 2), the polystyrene chains being cross-linked to one another by an occasional $-\text{CH}_2-$ group, joining a benzene group in one chain to one in another so as to maintain a loose three-dimensional gel-like structure. The interior of such a resin particle thus contains hydrocarbon residues and anionic sulfonate groups, which are fixed in the network structure. In the interstices of the network are water and solute molecules, which can move with considerable freedom in and out, and also sodium ions which balance the negative charge on the sulfonate groups; these have a fair degree of freedom of motion, but are of

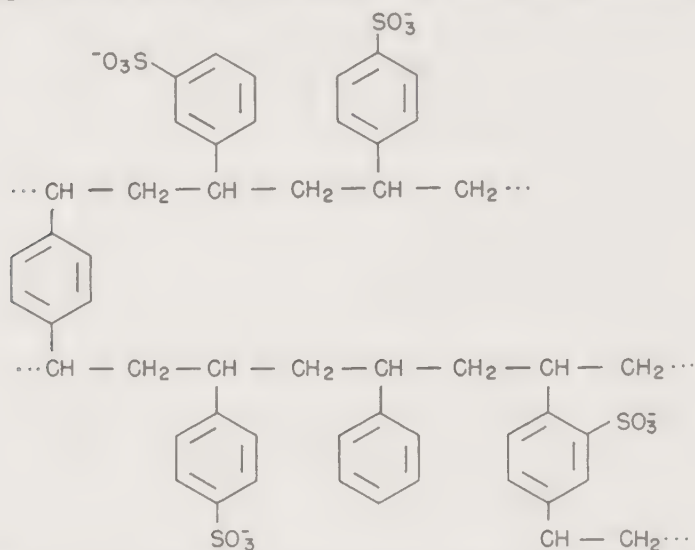


FIG. 2. Diagrammatic view of a polystyrene sulfonate cation exchange resin. The polystyrene chains are cross-linked here and there by divinylbenzene linkages, in which a benzene ring connects two aliphatic chains. The SO_3^- groups on the benzene rings require the presence of cations in the surrounding medium to maintain electrical neutrality. These may be simple cations, such as Na^+ or K^+ , or they may be larger cations, such as those of amino acids, peptides, or proteins. If there are many cross-links the structure is so tight that large cations cannot penetrate into the small interstices. For the fractionation of large peptides by ion exchange procedures, therefore, the amount of cross-linking should be small. The affinity of an organic cation for the resin depends on the Van der Waals attractions between the uncharged groups in the cation and the benzene rings of the polystyrene framework, as well as upon purely electrostatic forces.

course restrained by strong electrostatic attractions. The proportion of cross-links in the network is important; the larger the proportion of cross-links, the tighter the network, and the smaller the interstices into which solute molecules can penetrate. Amino acid molecules can penetrate even into rather highly cross-linked resins, but larger molecules such as peptides require a more loosely cross-linked structure if they are to penetrate effectively into the interior of the resin, and exchange back and forth with the surrounding liquid.

The resin, in the form of small beads, is placed in a column about 100 cm in length, and sodium hydroxide solution, followed by buffer solution, is passed through it. Then a small quantity of protein hydrolyzate (or any mixture of amino acids) is placed on top of the column. A suitable buffer solution then flows down through the column, and the amino acids are carried slowly along by the flow, each moving at a characteristic rate depending on its interaction with the charged groups and the hydrocarbon groups in the material of the column. The gel-like structure of the resin phase is important here; the evidence indicates that the amino acids can actually penetrate into the interior of the resin particles, and that they are passing constantly back and forth between the interior of the resin phase and the surrounding buffer solution. By suitable choice of the *pH* of the buffers, their composition, and their temperature, sharp resolution of the different amino acids into different portions of the effluent liquid can be obtained. These portions can be collected in a large series of small fractions, by means of a fraction collector; ninhydrin is added to each fraction, and the amount of amino acid present in each is determined by the intensity of the color developed. The identity of the amino acid in each portion is known from the sequence in the different portions of the outflowing liquid; the volume of liquid which must pass through the column before a given amino acid emerges is quite reproducible if the system is carefully defined. Recent developments have eliminated the necessity for collecting separate fractions; the effluent liquid meets and mingles with a stream of ninhydrin solution, the mixture flows through a heating coil to bring about the reaction that develops the characteristic blue color, and the concentration of amino acid is determined by passing light of wavelength 570 $m\mu$ across a section of the flowing liquid, and reading the light transmission with a photoelectric recording system.³

A sample of the results obtained by this procedure is shown in Fig. 3. The area under each section of the curve in the figure is proportional (with a few small correction factors for some amino acids) to the amount of the corresponding amino acid present. Such a complete analysis can be carried out on the hydrolyzate from 3 to 5 mg of protein. This is more than is required by paper chromatography, but the accuracy achieved is considerably greater, and the amount of material required is minute compared to that needed in most of the "classical"—and generally far less accurate—methods employed before the days of chromatography.

³ Proline and hydroxyproline react with ninhydrin to give a different product, with a maximum light absorption at 440 $m\mu$. A second photoelectric detector can be set to read the transmission at this wavelength, and thus to calculate the proline and hydroxyproline present.

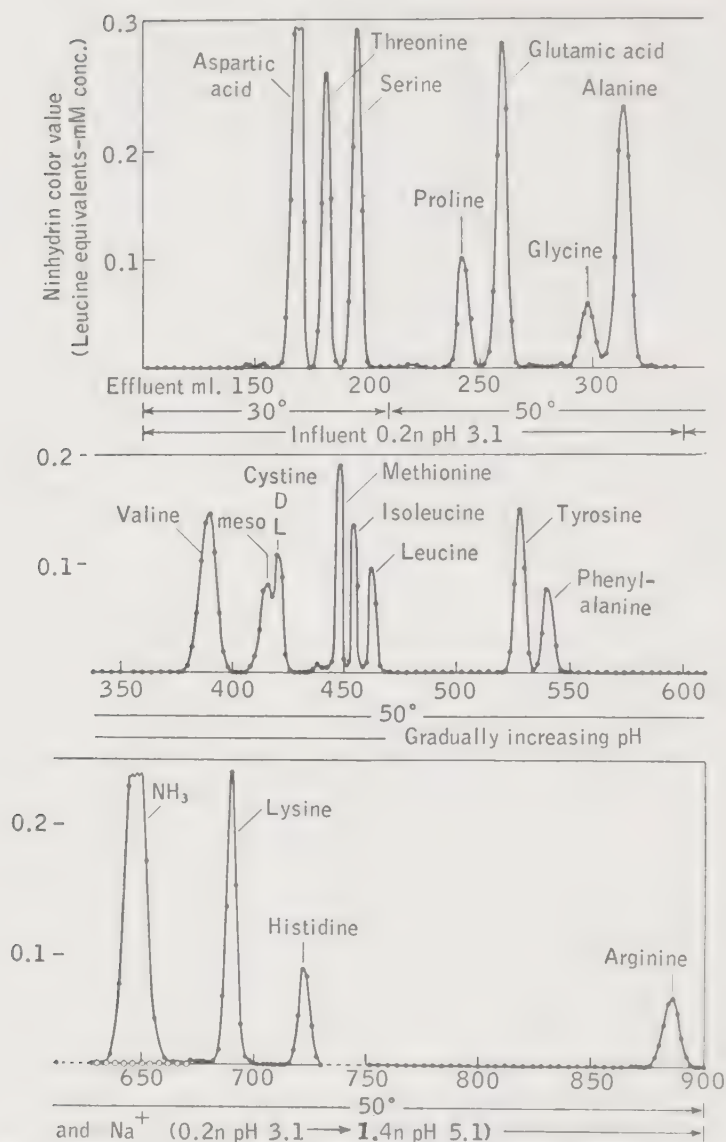


FIG. 3. The amino acids in a 70-hour hydrolyzate of ribonuclease A. Chromatography was carried out on a 150×0.9 cm. column of Dowex 50-X4. (From Hirs, Stein and Moore (1954), *J. Biol. Chem.* **211**, 941.)

A totally different method of analysis for amino acids is the microbiological technique (see, for instance, Snell, 1946), in which the growth of a microorganism is measured in a medium which contains all the substances needed by the organism except one particular amino acid. Addition of this amino acid then promotes growth, to an extent which can be measured by determining the turbidity of the medium or (in some cases) the amount of acid liberated by the organism during growth. Addition

of a complex mixture, such as a protein hydrolyzate, usually affects growth only through the content of the particular amino acid which is deficient in the medium. As may be inferred from this brief description, the method is full of pitfalls, both chemical and biological, which we shall not attempt to discuss here. Nevertheless, in the hands of careful workers it has given many data of surprising accuracy, agreeing to ± 5 to 10% with the data obtained by chromatographic methods. The latter, however, are increasingly favored by analytical protein chemists.

A valuable survey of the amino acid analysis of proteins, and of many of the results obtained, has been given by Tristram (1953). A number of important methods, which we have not attempted to mention here, are there discussed, with further references.

Some of the recent data on the amino acid analysis of proteins are given in Table II. Many other proteins have been studied, and the group selected here represents only a small fraction of the known proteins which have been analyzed. These, however, are sufficient to illustrate something of the great variety which exists among the proteins with respect to the relative number of different kinds of amino acid residues. Some proteins, such as insulin, contain relatively large numbers of nonpolar and aromatic side chains, and not very many electrically charged side chains. Others, such as myosin, contain a very large proportion of positively and negatively charged amino acids. It should be remembered that the total sum of the negative charges near pH 7 is equal to the amount of aspartic acid plus glutamic acid, minus the amide nitrogen, which appears as ammonia on hydrolysis. To this figure must be added the number of free terminal carboxyl groups at the ends of peptide chains, if this number is known. The total number of positive charges is equal to the sum of the arginine plus lysine residues, plus those histidine imidazole groups which carry a positive charge at pH 7, plus any terminal α -amino groups at the ends of peptide chains. Myosin carries the largest proportion of such positive and negative charges, relative to the total number of residues, of any of the proteins listed in Table II.

Some proteins have a great excess of the positively charged over the negatively charged groups. This is notably true, for instance, of lysozyme and ribonuclease in Table II, and this means that lysozyme has a very alkaline isoelectric point (Chapter 9). Ovalbumin and β -lactoglobulin have a considerably larger proportion of free carboxyl groups, and their isoelectric points lie in the acid range. Horse myoglobin and horse hemoglobin have a fairly even balance between the two types of groups and have neutral isoelectric points.

TABLE II
AMINO ACID COMPOSITION OF SOME IMPORTANT PROTEINS

	Beef pancreas Beef insulin	Hen's egg lysozyme	β -Lacto- globulin	Hen's egg oval- bumin	Horse heart myo- globin	Horse hemo- globin	Bovine serum albumin	Rabbit myosin	Wool keratin	Collagen
Per cent nitrogen	15.9	17.5	18.6	15.6	15.76	16.9	16.8	16.07	16.3	18.6
Per cent sulfur	3.36	2.87	2.53	1.6	1.61	0.37	0.48	1.92	3.3	
Assumed molecular weight	5,734	13,683	15,000	37,000	45,000	16,700	68,000	69,000	(850,000)	

	Residues per mole protein					Residues per 10 ⁵ g.				
Alanine	3	12	10-12	30	35	15	54	48	73.0	46.4
Glycine	4	3	11-12	7	19	13	48	17	25.3	87.0
Valine	5	9	6	18	28	6	50	35	22.1	39.7
Leucine	6	2	8	44	32	{ 22	75	65	{ 110.0	86.3
Isoleucine	1	3	6	17	25		0	14		42
Proline	1	4	2	17	14	5	22	29	16.7	82.6
Phenylalanine	3	3	3	9	21	5	30	24	26.2	22.1
Tyrosine	4	6	3	8	9	2	11	19	18.8	25.7
Tryptophan	0	0	8	3	3	2	5	2	3.9	8.8
Serine	3	15	9	14	36	6	35	28	41.2	95.4
Threonine	1	10	7	16	16	7	24	29	42.9	53.9
Cystine/2	6	8	10	7	2	0	2.5	36	11.7	98.9
Cysteine	0	0	—	3	5	0	3	1	—	0
Methionine	0	4	2	8	16	2	4.5	4	22.8	4.7
Arginine	1	4	11	6	15	2	14	23	42.3	59.7

TABLE II (Continued)

	Beef insulin	Beef pancreas ribo- nuclease	Hen's egg lysozyme	β -Lacto- globulin	Hen's egg oval- bumin	Horse heart myo- globin	Horse hemo- globin	Bovine serum albumin	Rabbit myosin	Wool keratin	Collagen
Histidine	2	4	1	4	7	9	36	18	15.5	6.8	5
Lysine	1	10	6	29	20	18	38	60	81.4	18.9	30
Aspartic Acid	3	15	20	32	32	10	51	65	66.9	54.1	47
Glutamic Acid	7	12	5	48	52	19	38	78	150.3	96.0	77
Amide N	(6)	(16-17)	(18)	(28)	(33)	(8)	(36)	(39)	(85.7)	(83.2)	(47)
Hydroxyproline											107
Hydroxylysine											7
Total	51	124	128-131	320	387	143	541	588	780	948.5	1073

Data for beef insulin are taken from E. J. Harfenist, *J. Am. Chem. Soc.* **75**, 5528 (1953); other data are taken from extensive tables compiled by G. R. Tristram in "The Proteins" (H. Neurath and K. Bailey, eds.), Vol. IA, Chapter 3, Academic Press, New York, 1953. Those for beef pancreas ribonuclease are from C. H. W. Hirs, W. H. Stein, and S. Moore, *J. Biol. Chem.* **211**, 941 (1954), as revised by the same authors, *ibid.* **219**, 623 (1956); also Hirs, *ibid.* **219**, 611 (1956). Extensive and accurate data for normal and abnormal human hemoglobins are given by Stein *et al.* (1957).

Note that the values listed for beef insulin and beef pancreas ribonuclease are presumably exact, since the values given here have been verified by the complete sequence of amino acid residues determined for these proteins and given later in this chapter. Data for the other proteins are less certain. The probable error in most of the values given is at least 5% and in many cases 10% or even more.

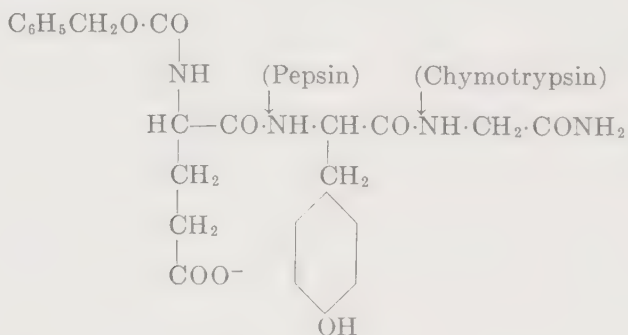
Most of the values for the amino acid content are given in moles of residue per mole protein. Wool keratin and collagen, however, are fibrous proteins for which no definite molecular weight can be given; and the molecular weight of rabbit myosin is subject to some uncertainty. The values for these three proteins, therefore, are listed as moles of residue referred to an arbitrary unit weight of protein of 10^6 g.

Polypeptide Chains and Their Presence in Proteins

All the existing evidence is compatible with the view that the amino acids in proteins are joined together by peptide linkages:

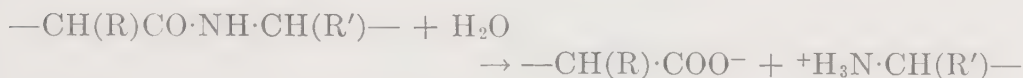


The evidence comes from many sources and is discussed in most books on organic chemistry and biochemistry. The most decisive evidence comes from the action of proteolytic enzymes; such enzymes have been studied in their action on many synthetic substrates, especially peptides and derivatives of peptides. Invariably these enzymes are found to split peptide bonds, and only peptide bonds, in all substrates tested (except for the special case of terminal amide and ester groups, discussed below). The particular bonds attacked depend on the nature of the R groups in the amino acids present, and to some extent on the nature of the neighboring groups. Thus carbobenzoxy-L-glutamyl-L-tyrosylglycyl amide:

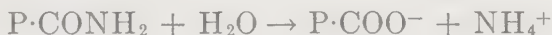


is split by both the enzymes pepsin and chymotrypsin. Both attack peptide bonds, but the particular bonds broken are different in the two cases, as indicated in the diagram. Both attack bonds involving tyrosine (or phenylalanine), but pepsin attacks the peptide linkage involving the NH groups of the tyrosyl residue, whereas chymotrypsin acts on the bond involving the CO⁻ group of this residue. It should be noted that, in this substrate, there is no terminal free amino or carboxyl group; the former is blocked by a carbobenzoxy group, the latter by the formation of an amide group. This is essential for the action of these particular enzymes; they do not attack peptide linkages adjoining free carboxyl or amino groups. Carboxypeptidase, on the other hand, attacks only a peptide linkage adjoining a free carboxyl group at the end of a peptide chain; aminopeptidases attack only a peptide linkage adjoining a free terminal

amino group. We shall not elaborate further on this important subject here; the essential point for our purpose is that, in every case, the attack is on a peptide linkage, and that the same enzymes also digest proteins. Moreover, when proteins are digested by these enzymes, acidic groups (carboxyl groups) and basic groups (amino groups) are liberated in equivalent amounts at every stage of the process, which is just what is to be expected if the bonds broken are peptide linkages:



One small qualification to the statement concerning enzyme specificity should be noted. Certain enzymes, such as trypsin, attack terminal amide bonds, if these are present at the ends of peptide chains, catalyzing the reaction

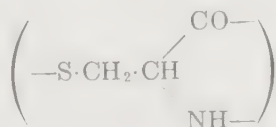


provided that the terminal group of the peptide chain P contains an arginyl or lysyl residue, for which trypsin is specific. These enzymes have been shown by Neurath and his associates to hydrolyze also the corresponding alkyl esters:

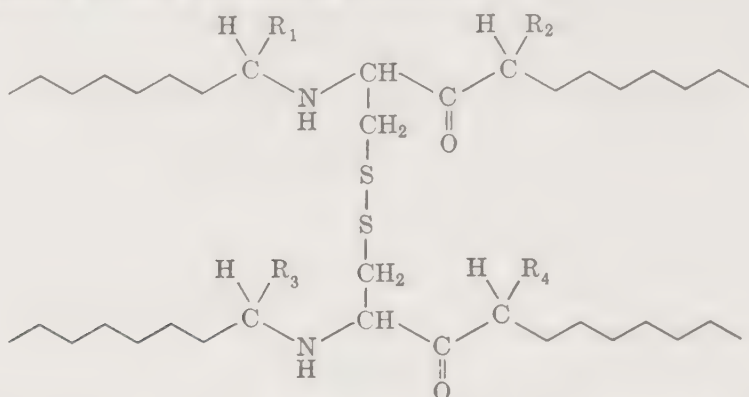


Indeed the esters are often attacked more rapidly than the corresponding amides. No evidence of such ester linkages in natural proteins has yet been forthcoming, however, so that this reaction, in spite of its great intrinsic interest, is irrelevant to the problem of protein structure. A detailed discussion of the specificity of proteolytic enzymes is given by Green and Neurath (1954).

Apart from the peptide linkage, one type of covalent bond appears to play a major part in protein structure—the disulfide bond of cystine. It can serve as a cross-link between the adjoining peptide chains, or between different portions of the same chain, provided that the latter loops back upon itself so as to bring two half-cystine residues close enough



together to permit disulfide bond formation. The importance of these cross-links will appear further as we proceed.



The peptide chains found in proteins can be relatively short—one of the two chains in the insulin molecule, for instance, contains only 21 residues. On the other hand, they may be many hundreds of residues long; human and bovine serum albumin, for example, are molecules with molecular weights near 66,000 containing 550 or more residues each; yet the best available evidence indicates that these residues are linked in a single long peptide chain, which is internally cross-linked in a number of places by disulfide bonds.

The complete chemical formula of a polypeptide chain is complex, even if the chain is relatively short. The usual system of naming chemical compounds becomes cumbrous when referring to such complicated structures. Therefore, we shall employ a shorthand notation, due to E. Brand, which is now widely used, referring to each amino acid residue by the first three letters of its name, as Ala for an alanyl residue, Glu for a glutamic acid residue, and so forth. The complete set of abbreviations is given in Table I in which the R groups are listed. The asparagine and glutamine residues, which contain amide groups attached to the β - and γ -carbon atoms, respectively, are denoted by the symbols Asp-NH₂ and Glu-NH₂. A half-cystine residue is denoted by CyS, an entire cystine by CySSCy, a cysteine by CySH. The convention is universally accepted that the sequence of residues, as written, begins at the left with the residue bearing a free α -amino group and ends with the residue bearing a free α -carboxyl group. The former is called by F. Sanger the N-terminal residue, and by C. Fromageot the initial residue; the latter is denoted as the C-terminal, or final, residue. We may illustrate by writing the formula of a sample heptapeptide, which in the usual chemical terminology would be denoted as L-lysyl-L-isoleucyl-L-tyrosyl-glycyl-L-cysteinyl-L-glutamyl-L-asparagine. (This particular peptide, to our knowledge, has not yet been synthesized or isolated; it is chosen simply for purposes of

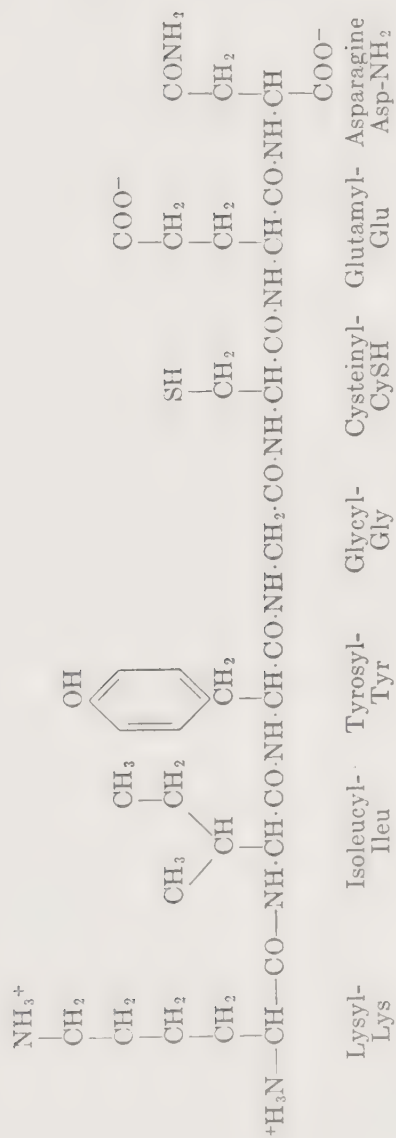
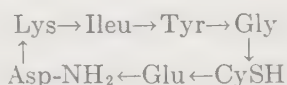
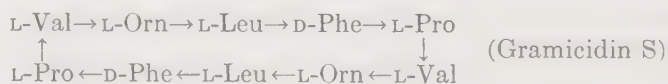


illustration.) All the optically active residues in this peptide are assumed, as in proteins, to be of the L-configuration; in the abbreviated formulas, we shall generally assume this to be the case and write the short formula for this peptide as Lys-Ileu-Tyr-Gly-CySH-Glu-Asp-NH₂. The complete chemical formula is given below, the ionizable groups being written with a positive or negative charge if this charged condition corresponds to their usual state at pH 7. This point is discussed in more detail in Chapter 8.

The end groups—for instance, the α -amino group of the N-terminal lysyl residue and the α -carboxyl group of the C-terminal asparagine in the peptide shown above—might of course form a peptide linkage, leading to the formation of a cyclic peptide, which in this case could be written



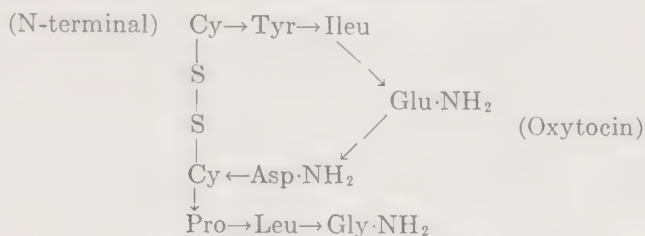
The arrow denotes the orientation of the CO—NH linkage in a cyclic peptide. In such a structure, of course, there would be no N-terminal or C-terminal group, although the side chains contain one free amino group (Lys), one carboxyl (Glu), one phenolic hydroxyl (Tyr), and one sulfhydryl (CySH). Certain cyclic peptides have been found in nature—for instance, the antibiotic Gramicidin S, obtained from *Bacillus brevis*, which apparently has the formula



Gramicidin S has at least two characteristics, apart from its cyclic structure, which set it apart from the peptides derived from proteins. One is the presence of phenylalanine in the D-configuration, the other the presence of the amino acid ornithine [$^+\text{H}_3\text{N}(\text{CH}_2)_3\cdot\text{CH}(\text{NH}_3^+)\cdot\text{COO}^-$]. Both these features—the presence of amino acids of the D-configuration, and of amino acids not found in proteins—are characteristic of many naturally occurring peptides, especially those derived from certain microorganisms. Consideration of such substances would lead us into a different realm, apart from that of protein chemistry; a discussion of many compounds of this class is given, for instance, by Bricas and Fromageot (1953).

Peptides may assume a cyclic form due to the formation of a disulfide bridge between two half-cystine residues, separated from one another by several intervening residues of the chain. In this case, one or both of the half-cystine residues may continue in another peptide chain projecting beyond the ring. Outstanding examples are certain hormones of the pos-

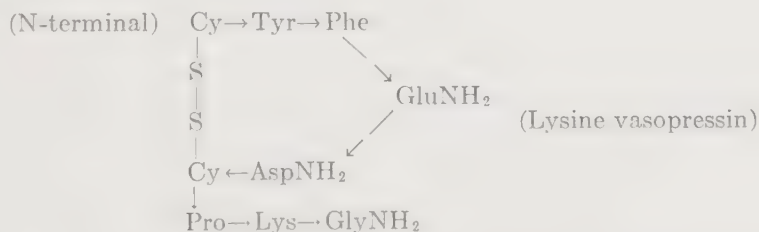
terior pituitary—oxytocin, which induces uterine contraction, and vasopressin, which influences blood pressure. These have been synthesized by du Vigneaud and his colleagues (1954, 1956), who have shown the synthetic products to have the full hormone activity of the same substances isolated from the pituitary. The amino acids which they contain are all among those found in proteins, and are all of the L-configuration. In abbreviated notation, the formula of oxytocin is:



The half-cystine at the upper left in the formula contains a free N-terminal amino group; there is no C-terminal carboxyl, since the carboxyl group of the terminal glycine residue exists as an amide ($-\text{NH} \cdot \text{CH}_2 \cdot \text{CONH}_2$).

The ring structure containing the disulfide linkage of oxytocin contains 20 atoms—8 of these involve the two half-cystine residues ($\text{C}(\text{O}) \cdot \text{CH} \cdot \text{CH}_2 \cdot \text{S} \cdot \text{S} \cdot \text{CH}_2 \text{CH} \cdot \text{NH}-$), and the other 12 are derived from the other four amino acid residues, each of which contributes 3 atoms to the ring. The ring structure is indicated more explicitly in Fig. 4, in which, however, no attempt is made to show the actual spatial relations involved. These relations are indeed not known; but the same ring pattern is found in the vasopressins and in one segment of the molecule of insulin (as we shall see later).

Two types of vasopressin have been identified by du Vigneaud; the structures of the two are very closely related to that of oxytocin, and to one another. Lysine vasopressin differs from oxytocin only in the substitution of a lysyl for the leucyl, and of a phenylalanyl for the isoleucyl group:



Arginine vasopressin differs from lysine vasopressin only in that arginyl replaces lysyl in the structure shown above.

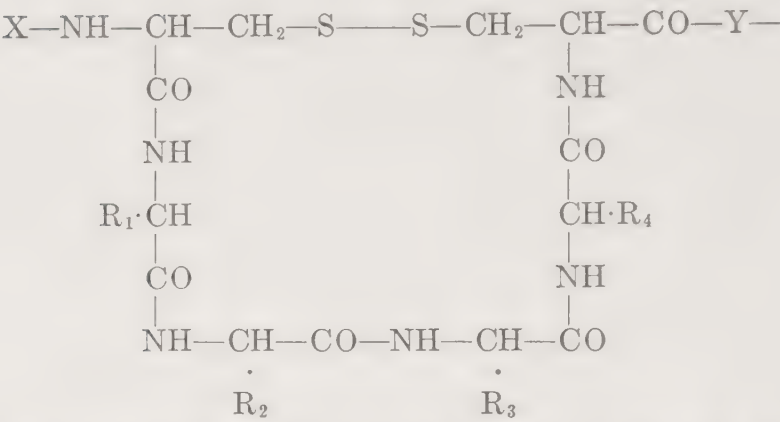


FIG. 4. The peptide-disulfide ring structure of oxytocin, vasopressin, and insulin. The attached group at positions X, Y, R₁, R₂, R₃, and R₄ are as indicated:

Substance	X	R ₁	R ₂	R ₃	R ₄	Y
Beef insulin	Val·Glu·Glu·NH ₂	CyS-	Ala	Ser	Val	Ser·Leu·Tyr . . .
Pig insulin	Val·Glu·Glu·NH ₂	CyS-	Thr	Ser	Ileu	Ser·Leu·Tyr . . .
Sheep insulin	Val·Glu·Glu·NH ₂	CyS-	Ala	Gly	Val	Ser·Leu·Tyr . . .
Oxytocin	H	Tyr	Ileu	Glu·NH ₂	Asp·NH ₂	Pro·Leu·Gly NH ₂
Arginine vasopressin	H	Tyr	Phe	Glu·NH ₂	Asp·NH ₂	Pro·Arg·Gly NH ₂
Lysine vasopressin	H	Tyr	Phe	Glu·NH ₂	Asp·NH ₂	Pro·Lys·Gly NH ₂

Determination of Arrangement of Peptide Chains and Their Linkage within Protein Molecules

When a purified protein is obtained, its minimum molecular weight may be determined by analysis. In special cases, for instance the hemoglobins, the determination of a metallic or other constituent which forms an integral part of the molecule may give the desired information. Most mammalian hemoglobins contain very nearly 0.335% iron. Since the atomic weight of iron is 55.85, the smallest weight of hemoglobin which can contain one atom of iron is $55.85 \times (100/0.335) = 16,700$. The true molecular weight may be any integral multiple of this minimum amount; in the case of the hemoglobins we know from other evidence that the correct multiple is 4. Likewise, analysis for any amino acid residue which forms only a small fraction of the total weight of the protein may be used to calculate a minimum molecular weight. The calculation of the actual molecular weight, however, depends on the methods of physical chemistry—osmotic pressure, light scattering, sedimentation equilibrium in the ultracentrifuge, sedimentation velocity combined with diffusion

measurements, and a variety of other techniques which will be discussed in Volume II. At present we must simply assume that evidence concerning molecular weight is available, from one or more of these methods, for the protein under study.

An amino acid analysis on the hydrolyzed protein, yielding such data as those of Table II, then permits the calculation of the total number of amino acid residues of each kind in the protein, although, especially if the molecule is large, these numbers may not be known to the nearest integer. In any case, such information falls a long way short of telling what we should like to know about the structure of the protein. Further knowledge must then be sought to answer the following questions:

1. How many peptide chains are present in the molecule? Are they open chains, or cyclic? If the chains are open, what is the nature of the N-terminal and of the C-terminal group in each chain?

2. If there is more than one chain present, how are the chains held together? Are the cross-linkages covalent, such as disulfide links of cystine or the phosphate cross-linkages described later in this chapter, or are they weaker links, owing perhaps to hydrogen bonding between polar side-chain groups of different peptide chains?

3. Are there internal cross-links, involving the formation of loops within a single peptide chain, produced for instance by the joining of two half-cystine residues to form a disulfide bridge between different portions of the chain?

4. What is the sequence of amino acid residues in each peptide chain of the molecule?

5. What is the detailed arrangement in space of the whole structure? Is the structure rigid, or is the molecule flexible, at least in part?

6. If the molecule possesses specific biological activity, such as that of a hormone or an enzyme, how is this activity correlated with its structure?

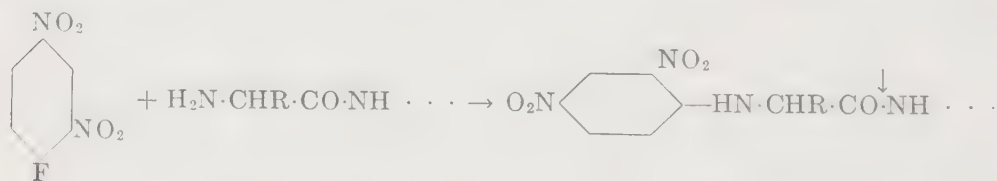
It is obvious that a complete answer to all these questions would be a very difficult achievement indeed. In the case of three species of insulin, however, the brilliant work of F. Sanger (Sanger *et al.*, 1951-55) has answered all these questions except the last two. The fifth question—the detailed arrangement of the molecule in space—is most likely to be answered by X-ray diffraction studies on the crystalline protein. Although such studies have made great progress in recent years, no structure of such a complex molecule has yet been worked out in detail. For some of the fibrous proteins, however, much is now known concerning the repeating patterns of the polypeptide chains—notably in silk fibroin, keratin, and collagen. We shall return to them later.

The sixth question, concerning the biological significance of the chem-

ical structure, is of course the most fundamental of all; the desire to answer it is probably the chief motive which has led investigators to undertake the arduous task of unraveling such complex structures. Certainly the establishment of a rational and detailed correlation between structure and biological function of proteins will have a revolutionary influence on the biological and medical sciences. Probably such understanding will be achieved with certain enzymes before it is achieved for hormones, for the enzyme-substrate system can be studied *in vitro*, whereas the actions of hormones are far more complex. There is as yet no case, however, in which knowledge has progressed far enough to permit a discussion of this question, based on detailed structural evidence. Therefore we merely call the attention of the reader to the supreme importance of the problem, and return now to the first four basic structural questions posed above.

The Number of Peptide Chains in the Molecule, and the Nature of the Terminal Groups

This problem has been approached by attaching a labeling group to one kind of terminal group—a group which still remains attached while the protein is hydrolyzed to its constituent amino acids. The label most widely employed is the dinitrophenyl group, obtained by the reaction of 2,4-dinitrofluorobenzene (DNFB) with terminal amino groups:



The linkage forms with amino groups in mildly alkaline solution, under conditions in which most proteins are not likely to suffer extensive denaturation and certainly not to undergo hydrolysis. Subsequent hydrolysis of the peptide linkages in acid solution (see the vertical arrow in the diagram) can be carried out so as to give the 2,4-dinitrophenyl amino acid containing the R group which was present at the amino end of the peptide chain. This derivative, commonly denoted for brevity as a DNP-amino acid, is strongly yellow in color; its color, and its characteristic speed of migration on a paper chromatogram, permit its identification, by comparison with a set of known DNP derivatives, prepared from pure amino acids. Unfortunately some breakdown—often 20 or 30%—of the DNP-amino group linkages occurs during hydrolysis, and this must be corrected for. Since the number of N-terminal groups per protein molecule is usually small, however, this number can generally be determined

to the nearest integer. This method, developed by Sanger in 1945, has proved a tool of extraordinary power in determining the amino end groups of proteins, and also—as we shall see below—in determining the sequences of amino acids in the peptide chains, by labeling the N-terminal groups obtained by partial hydrolysis of the long chains.

We note that the free ϵ -amino groups of the lysine residues in the protein may also form DNP derivatives, giving ϵ -DNP-lysine in the hydrolyzate. This, however, can be separated chromatographically from the N-terminal α -DNP amino acids. A lysyl residue in the N-terminal position, as in ribonuclease or lysozyme, gives α, ϵ -di-DNP-lysine on hydrolysis. O-DNP derivatives of tyrosine may also be formed by the reaction of DNFB with proteins and peptides.

Another powerful method, developed by Edman, employs phenylisothiocyanate (C_6H_5NCS), which reacts with the terminal amino groups in solvents such as dry pyridine, followed by treatment with anhydrous hydrochloric acid. As a result, a phenylthiohydantoin, containing the characteristic R group of the N-terminal amino acid, is split off the end of the chain. The amino group of the next amino acid in the sequence of the peptide chain is left as a free N-terminal amino group. The treatment can then be repeated, and the successive amino acid residues “peeled off” one by one, from the N-terminal end, and identified. The process, described in this simple fashion, might appear ideal for the determination of structure. In practice, unfortunately, it generally gives trouble after the first four to six residues have been taken off; losses of material in the successive steps, and the occurrence of side reactions, generally cause the method to become ineffective and to yield ambiguous results long before it has been possible to work down the entire length of a long chain. The phenylthiohydantoin method, therefore, although it has given valuable results, has proved hitherto of more limited value than the DNP method (see Sanger, 1952; Anfinsen and Redfield, 1956).

Methods for the determination of the C-terminal amino acids are in general less satisfactory than those for the N-terminal acids. The one most widely used involves treatment of the protein with the enzyme carboxypeptidase, derived from pancreas. Most C-terminal amino acids are split off from peptide chains by this enzyme; it attacks most readily residues with aromatic side chains; those with strongly polar side chains are less rapidly attacked; and a prolyl residue immediately adjoining the C-terminal residue completely halts the progress of the reaction. When the enzyme has removed the C-terminal group, it then attacks the next group, which is now C-terminal, and continues until blocked by the presence of an adjoining prolyl residue. Thus it is necessary to make quantitative determinations of the amounts of the various amino acids re-

leased, as a function of time, and to interpret the results carefully. Especially in proteins containing two or more peptide chains, the results may be ambiguous.

It is not our purpose here to go into the details of these and other methods which have been used for end group determinations. Extensive

TABLE III. N-TERMINAL AND C-TERMINAL RESIDUES

Protein	Assumed molecular weight $\times 10^{-3}$	N-Terminal residues	C-Terminal residues
Aldolase (rabbit)	140	2 Pro	
Carboxypeptidase	34	1 Asp·NH ₂	
Pepsin*	35	1 Leu	
Trypsinogen (cattle)	20	1 Val	
Trypsin (cattle)	20	1 Ileu	1 Lys
Pancreatic trypsin inhibitor	9	1 Arg	
Soybean trypsin inhibitor	24	1 Asp	1 Leu
Chymotrypsinogen	21.5	None	
α -Chymotrypsin	21.5	1 Ileu, 1 Ala	1 Tyr, 1 Leu
δ -Chymotrypsin	21.5	1 Ileu	1 Leu
Hemoglobin (horse)*	66	6 Val	
Hemoglobin (cattle, sheep, goat)*	66	2 Val, 2 Met	
Hemoglobin (human)*	66	5 Val	
Serum albumin (human, cattle)*	66	1 Asp	
Lysozyme (egg white)*	44	1 Lys	1 Leu
Ovalbumin (egg white)*	45	None	1 Pro
Conalbumin (egg white)	87	1 Ala	
Growth hormone (somato- tropin)*	45	1 Phe, 1 Ala	1 Phe

Most of these data are taken from the paper by F. Sanger in "Fibrous Proteins and Their Biological Significance" (1955)—see under General References at the end of the chapter. For data on insulins, corticotropins, glucagon, and melanophore-stimulating hormones, see Table IV. For ribonuclease, see Fig. 6.

* Denotes a protein for which further details regarding amino acid sequences are given by Anfinsen and Redfield (1956). Rhinesmith *et al.* (1957) have recently reported only 4 N-terminal valine residues in human hemoglobin for an assumed molecular weight of 66,000, using the DNP method, in a very careful study. These four chains fall into two different classes; in two of them the N-terminal sequence is Val-Leu-; in the other two the residue adjoining the N-terminal valine has not yet been determined, but is certainly different.

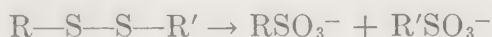
and critical discussions are given by Sanger (1952) and by Anfinsen and Redfield (1956), with abundant references to the original sources. Here we give only a sample of some of the results obtained, on a few important proteins (Table III).

It may be seen from Table III that almost any amino acid may serve

as the N-terminal group of a peptide chain in a protein—acidic, basic, hydroxylic, and nonpolar side chains all may occupy the end position. The same is probably true of the C-terminal groups, though here the data are far less extensive.

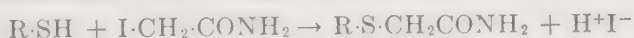
Cross-Linkages Between Peptide Chains: Disulfide and Phosphate Cross-Linkages

The manner in which a disulfide link of cystine can cross-link two peptide chains—or two portions of the same chain—has already been illustrated (p. 76). To determine which type of link is involved, the disulfide bonds must be broken. This can be done either by oxidation or by reduction. Sanger has used oxidation with performic acid (prepared by mixing formic acid and hydrogen peroxide) to convert disulfide links of cystine to cysteic acid groups:



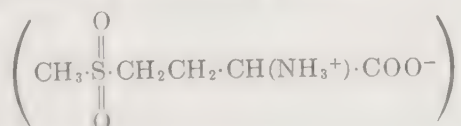
If a protein is made up of a single peptide chain, and the disulfide links serve merely to cross-link different parts of the chain into loops, such oxidation should not cause any significant change in molecular weight, apart from the small change involved in the conversion of each S—S linkage to two —SO₃[−] groups.

Performic acid oxidizes the indole ring of tryptophan, but it is not known whether this would lead to any breakage of the peptide chain at the tryptophan residue. Insulin and ribonuclease do not contain tryptophan, and these are the two proteins for which performic acid oxidation has already been employed with conspicuous success for the elucidation of structure. In other cases it will have to be determined whether it will be preferable not to oxidize the disulfide bonds, but to break them by reducing them to —SH groups, using some reagent such as thioglycolic acid. The —SH groups, once formed, are very reactive, and might form new disulfide cross-links by oxidation in air. Such reactions, however, can be blocked by alkylating the —SH groups immediately with a reagent such as iodoacetamide:



The number of —SH groups thus reacting can be determined by measuring the additional number of amide groups introduced into the protein, as determined by the ammonia released on acid or alkaline hydrolysis. This method has been used by Dr. Margaret Hunter at Harvard Medical School in the study of the disulfide linkages of serum albumin.

Methionine is also oxidized to methionine sulfone



by performic acid. This does not, however, essentially disturb the analysis for methionine in a protein treated with performic acid. In the resulting hydrolyzate one must of course analyze for methionine sulfone, rather than methionine itself.

When beef insulin was oxidized with performic acid, the molecule was broken up into two kinds of peptide chains (Sanger 1952, 1956), one predominantly acidic (the A chain), with 21 amino acid residues, including four cysteic acid groups; the other predominantly basic (the B chain), with 30 amino acid residues, of which two were cysteic acid groups. It was evident that the two kinds of chain were linked by disulfide bonds in the original insulin molecule, and that oxidation had broken the links. There was no evidence of the presence of other peptide chains (cyclic chains, for instance) among the oxidation products, and the A and B chains were present in equimolar proportions. Moreover the most careful determinations of the molecular weight of insulin had indicated a value near 6000 for the fundamental molecule—a value which accorded with the view that the molecule consisted of one A and one B chain, linked together by disulfide bonds.⁴

Beef pancreas ribonuclease, although the fundamental unit is more than twice as large as in the case of insulin—124 amino acid residues as against 51 for insulin—consists only of a single peptide chain, which, however, is cross-linked by four disulfide linkages. Performic acid oxidation therefore leads to no significant change in molecular weight, although the oxidized product is a looser and more random structure than the original molecule and is devoid of enzymatic activity (see Anfinsen and Redfield, 1956, for extensive discussion and references). Even the much larger molecules of human and bovine serum albumin—molecular weight near 66,000, with nearly 560 amino acid residues—appear to be made up of a single peptide chain, with 17 disulfide cross-links hooking different parts of the chain together (see, for instance, Low and Edsall, 1956).

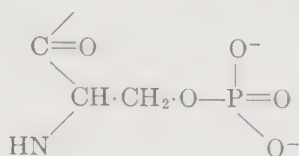
The very important fibrous protein keratin contains large quantities of cystine residues, and the cross-links which these provide between the peptide chains have a profound effect on the mechanical properties of keratin fibers. There is thus no doubt concerning the profound importance of disulfide cross-links for protein structure.

PHOSPHATE CROSS-LINKAGES

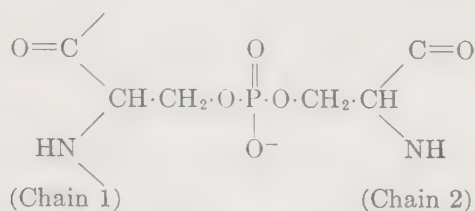
Cross-linkages through phosphate groups have been revealed as significant in several proteins, by the recent work of Perlmann (1955).

⁴ The determination of the molecular weight of the fundamental unit was a very difficult problem. The 6000 molecular weight unit polymerizes very readily; in acid aqueous solution, the smallest unit commonly found consists of two of these fundamental units (molecular weight, 12,000) and in neutral aqueous solution the molecule most commonly found has a molecular weight of 36,000 to 48,000. Osmotic pressure measurements in solutions of guanidine hydrochloride (Linderström-Lang), and countercurrent extraction with organic solvents (Harfenist and Craig) were necessary to obtain dissociation of insulin into the fundamental unit.

Phosphate groups may be attached in monoester linkage to one of the hydroxyamino acid residues in a peptide chain:



Of course, no cross-linkage is involved here. They may, however, be involved in a diester linkage between two such chains, or link different parts of the same chain together into a loop. The cross-linkage of two seryl residues may be given as an example:



Also the linkage may involve an amino or other nitrogenous group on an amino acid residue of one of the two chains, and a hydroxyl group of the other. Such a linkage may be called an —N—P—O— bond, in distinction from the —O—P—O— bond illustrated above. Finally the cross-linkage may involve a pyrophosphate, instead of an orthophosphate, bridge between two chains or chain segments; such a bridge may be designated as an —O—P—O—P—O— cross-link. All these linkages have been found by Perlmann in proteins, by employing different phosphatases specific for each type of linkage in smaller molecules. Thus in the milk protein α -casein, 40% of the phosphorus is present in the form of simple primary phosphate esters, linked presumably to hydroxyamino acids. Another 40% is held in diester linkages; the specificity of the enzyme used indicates that these linkages are of the —N—P—O— type discussed above. Finally 20% of the phosphorus atoms are involved in a pyrophosphate linkage of the type —O—P—O—P—O— . The splitting of this linkage by pyrophosphatase involves no release of inorganic phosphate into the solution, but such release can be produced by subsequent addition of the prostate phosphatase which splits primary phosphate esters. Thus, it appears that this pyrophosphate linkage actually serves as a cross-link between peptide chains. Breakage of these cross-links results in a separation of the protein molecule into smaller units, indicating that these phosphorus linkages actually hold different peptide chains together. In β -casein, likewise, diester cross-linkages are found, but here

the specificity of the enzyme involved indicates that the linkages are of the type —O—P—O— , illustrated above. Here again it is found that, after removal of the phosphorus, material soluble in trichloroacetic acid is present in the solution, so that a splitting into smaller molecules has occurred, as a result of the breakage of the —O—P—O— linkage. Finally Perlmann has pointed out that the single atom of phosphorus present in the pepsin molecule is probably also present in an ester linkage of the O—P—O type. Pepsin appears to be a protein consisting of only one peptide chain; in this case, therefore, the phosphorus must cross-link a portion of the chain into a cyclic loop. It is still uncertain how general the significance of these phosphate cross-linkages in proteins may prove to be. They are certainly much less common than the disulfide cross-linkages, but their discovery is an important event in enlarging our general picture of the possibilities of protein structure.

Other Cross-Linkages

As yet we have spoken only of cross-linkages involving covalent bonds. Cross-links involving hydrogen bonding between amino acid side chains, or perhaps between groups in the side chains and the CO or NH bonds of the peptide linkages, may also be of great importance in fixing the protein molecule in its native state. A protein molecule may hold together in solution as a functional unit with a high degree of stability, even when it consists of several different peptide chains with no covalent cross-links between them. This, for instance, appears to be the case with horse hemoglobin, for which Porter and Sanger showed the presence of six N-terminal amino groups per molecule (see Table III), and, therefore, by inference, of six different peptide chains. Chemical studies by Ingram (1955), and by Benesch *et al.* (1955) on the sulfhydryl groups of horse hemoglobin have shown that six sulfhydryl groups are present, but no disulfide cross-links. Therefore, since there is also no evidence whatever of phosphate cross-linkages in hemoglobin, other types of bonds must be invoked to explain the stability of the molecule. We shall not pursue the subject further here, since a more detailed discussion of hemoglobin will be given in a later chapter.

In some proteins such as ovalbumin and ribonuclease, the phenolic hydroxyl groups of tyrosine residues appear to be cross-linked with other groups, perhaps by hydrogen bonding with ionized carboxyl groups of aspartic or glutamic acid residues, of the type illustrated on p. 62. The evidence here is derived from studies of the ultraviolet absorption of tyrosine between 275 and 300 $\text{m}\mu$, which shifts to longer wavelengths and becomes more intense when the tyrosine hydroxyl group ionizes. This evidence is discussed in more detail in Chapter 9. How much strength these bonds actually contribute to the native structure of the protein is

still uncertain. It is quite possible that other bonds, as yet unidentified, are more important in maintaining the molecule in the native configuration, and that the breakage of the links involving the tyrosine hydroxyls may be a secondary results of the breakage of these other linkages. It seems likely, however, that the bonds involving the tyrosine hydroxyls are important in maintaining the native structure. We know in any case that there are other proteins in which this is not true; Crammer and Neuburger (1943) found the tyrosine groups in insulin to react as if they were part of a simple tyrosine peptide. Tanford and Roberts have found an intermediate state of affairs in serum albumin (see Chap. 9).

The cohesion between nonpolar side chains is probably often important as a stabilizing factor. As Kauzmann (1954) has pointed out, these side chains tend to hold together with considerable force. The energy required to separate them is decidedly greater than would be calculated from van der Waals attraction alone, for the separation of two nonpolar side chains involves bringing each of them in contact with additional water molecules and, therefore, requires the breaking of some hydrogen bonds between water molecules. The stabilization of serum albumin against denaturation by heat, or by urea and guanidine hydrochloride, by the addition of fatty acid anions with long nonpolar side chains is probably associated with a stabilizing interaction of this sort—the fatty acid anions certainly bind to the albumin, and they probably tie down some potentially loose and unstable points in the albumin structure.

Sequence of Amino Acid Residues in Peptide Chains

The determination of these sequences is almost entirely an achievement of the years since 1945, and especially since 1950. The first and, because of its pioneering character, the most important achievement in this field was the determination of the complete sequence of amino acid residues in the A and B chains of insulin by Sanger and his associates (see Sanger, 1952, 1956, for reviews). The method adopted involved breaking up each peptide chain, by partial hydrolysis, into a number of shorter peptide chains, separating them from one another by chromatography or by migration in an electric field, analyzing each of the shorter chains to determine its amino acid composition and generally to determine the amino acid sequence within it also, and finally fitting together all the data on the breakdown products to reconstruct the sequence in the original chain.

Two methods of partial hydrolysis were employed: (1) acid hydrolysis, which gives a more or less random splitting of the various peptide linkages; (2) hydrolysis with enzymes, in which only certain specific peptide linkages were split. The use of both methods was found to be essential, in order to obtain enough information to solve the problem.

The use of the first method may be illustrated by considering the determination of structure in (say) an octapeptide, which we may designate for brevity by the formula ABCDEFGH, each letter denoting an amino acid residue. Total hydrolysis of an aliquot of the material to amino acids, and analysis of the hydrolyzate, will demonstrate the presence of the constituent amino acids and the proportions in which they occur. Formation of a DNP-peptide, and subsequent hydrolysis on another aliquot, will demonstrate that A is the N-terminal group, and the action of carboxypeptidase may be used to show that H is the C-terminal group. If partial hydrolysis is now carried out—for instance by employing concentrated hydrochloric acid at 37° for two or three days—partial breakdown will occur, and a mixture of peptides will be obtained. These must of course be separated by fractionation, generally by paper chromatography (Turba, 1954; Thompson and Thompson, 1955) or by flowing down a column of an ion exchange resin (Moore and Stein, 1956), or by the application of an electric field (ionophoresis) if some of the peptides carry different net charges from others. A partially hydrolyzed mixture from such a peptide as ABCDEFGH will contain some of the original peptide, some of each of the free amino acids derivable from it on complete hydrolysis, and some of each of the possible intermediate peptides, such as AB, EF, BCD, DEF, DEFG. The relative yields of these different products depend on the relative time of hydrolysis, and on the relative susceptibility of the different bonds to splitting. At zero time, only the original peptide is present; at infinite time, only the free amino acids (neglecting any decomposition they may have undergone during the procedure). Thus an appropriate hydrolysis time must be chosen which will give suitable intermediates in the best possible yield.

Problem. We may illustrate by an example, say a hexapeptide ABCDEF. Assume for simplicity that all the five peptide bonds, A—B, B—C, C—D, D—E, and E—F, are attacked at the same rate, and that this rate is the same whether or not the other neighboring bonds are split or not. If hydrolysis has proceeded so that the extent of splitting of each of the five bonds is α ($0 \leq \alpha \leq 1$), what is the relative concentration of each of the possible products of hydrolysis?

Consider for instance the dipeptide A—B. The requirement for its existence in the hydrolyzate is that bond A—B is unsplit (the probability that this is true is, by our hypothesis, $1 - \alpha$) and that bond B—C is split (probability α). The probability that both these conditions are simultaneously fulfilled is equal to the product of the two separate probabilities. Hence the relative concentration of this peptide is $\alpha(1 - \alpha)$. For the occurrence of the peptide B—C the condition is that both the bonds A—B and C—D are split (combined probability α^2), but that B—C is unsplit (probability $1 - \alpha$). The relative concentration of B—C is thus $\alpha^2(1 - \alpha)$, and the same value holds for any other dipeptide in which neither amino group is terminal, on the assumption of equal rate of splitting for all bonds. Calculate in this way the concentrations of all the different possible peptides (including free amino acids) as a function of α .

What are the numerical values for all of them when $\alpha = 0.5$? How much of the original hexapeptide is still left at this stage?

Suppose you want some particular hydrolysis product—say the tripeptide CDE—for experimental purposes. How far should hydrolysis proceed so that you may get a maximum yield of it?

Note: The assumption of equal rates of splitting for all bonds, employed above, is unrealistic in practice. R. L. M. Synge, for example (see Sanger, 1952), studied the relative rates of hydrolysis of a number of dipeptides in strong acid. Taking the rate for Gly·Gly as unity, that for Gly·Leu was 0.40, Gly·Val 0.31, Leu·Gly 0.23, Leu·Leu 0.048, and Val·Gly only 0.015. Peptide linkages involving valine, especially if the valyl residue is on the carboxyl side of the linkage, appear to be especially resistant to hydrolysis. Moreover, the rate of splitting of any given bond is certainly not entirely independent of other bonds present near it in the same molecule. The actual situation is thus very complex, but the simple analysis in the problem given above is useful for grasping the general features of the actual situation.

If it is possible to obtain all the dipeptides AB, BC, CD, DE, EF, FG, and GH from the original octapeptide, and if there is good reason to believe that no new recombinations of peptide linkages have occurred during hydrolysis, then the existence of all these dipeptides (and of no others) would fix the original sequence uniquely. The partial hydrolyzate, of course, contains tripeptides, tetrapeptides, and all the higher peptides up to the original octapeptide, in varying amounts. The sequence in each of these may be determined, although the labor of doing so in the case of the larger peptides may be considerable, and each provides relevant information concerning the original structure. Thus the total amount of information potentially obtainable from all the breakdown products is far more than enough to determine the original sequence. This excess of information, however, is a very valuable thing; if all the data obtained fit together, and are consistent with one unique sequence in the original chain, then the evidence for the structure of that chain is greatly strengthened. On the other hand, if discrepancies appear, they might mean either that some new recombination of peptide bonds had occurred during hydrolysis, or that the original preparation was a mixture of two or more peptides with different sequences. In the case of insulin, Sanger's data could all be explained in terms of a single sequence for the A chain, although the approach outlined here proved, for various reasons, inadequate in itself to establish the sequences and had to be supplemented by the use of hydrolysis with enzymes.

The determination of a dipeptide structure is straightforward; one must identify the two amino acids produced on hydrolysis and determine which of the two forms a DNP-derivative when the peptide is treated with 2,4-dinitrofluorobenzene. For a tripeptide, the same procedure, followed by a determination of the C-terminal amino acid with carboxypeptidase, generally fixes the structure uniquely. For a tetrapeptide, the

total amino acid content, and the N- and C-terminal groups, may be fixed by the same procedure, but the position of the other two residues may be uncertain. For instance we might know that the peptide is either Asp·Glu·Ala·Gly or Asp·Ala·Glu·Gly. In this case we indicate the region of uncertainty by enclosing it in parentheses, writing Asp(Glu,Ala)Gly. The residues within the parentheses are known to be present, but their order is uncertain. If further experiment definitely fixes the order, we omit the parentheses and write (for instance) Asp·Glu·Ala·Gly. Obviously, the larger the peptide is to begin with, the more work is required to determine the structure completely. In fact, for very long peptides—say those containing 100 residues or more—such an approach, though it may yield valuable information, is never likely by itself to give a complete solution to the sequence problem. Since there are only about twenty different kinds of amino acids, the same ones, often in rather similar proportions, will turn up in breakdown products arising from quite different parts of the original chain. Moreover, in such a complex mixture, the adequate separation of the different components becomes a very difficult matter; one may still be dealing with a mixture of two or more peptides, after fractionation of the mixture by chromatography or ionophoresis, but may mistake it for a pure compound. In such cases, of course, the whole basis for the sequence determination goes astray. The actual procedure, even if the original peptide is of moderate size, requires great skill and judgment on the part of the experimenter, and the pitfalls are numerous.⁵

The second approach, employing enzymatic hydrolysis, was employed by Sanger as an essential part of the solution to the problem of the structure of insulin. It has since been developed further, especially by the investigators who have worked on the structure of ribonuclease (Hirs *et al.*, 1956; Redfield and Anfinsen, 1956; Anfinsen and Redfield, 1956). Each enzyme attacks only certain types of peptide linkages, so that the splitting of the original chain is specific, not random. The simplest proteolytic enzyme, with regard to the specificity of its action, is trypsin, which attacks only linkages containing a lysyl or arginyl residue on the carbonyl side of the CO·NH group; hence the products of tryptic hydrolysis are peptides with C-terminal lysine or arginine. Such hydrolysis leads to a relatively small number of peptides, with an average number

⁵ Another limitation of partial acid (or alkaline) hydrolysis is that peptide bonds involving the —NH— groups of seryl or threonyl residues are very labile. All such bonds were broken in the partial hydrolyzates prepared by Sanger from the A and B chains of insulin. Therefore the nature of the groups on the carboxyl side of these linkages could not be determined by these methods alone. Data obtained from the peptides obtained by enzymatic hydrolysis were necessary to settle this problem, among others.

of residues of the order of 10 per peptide in ribonuclease. Redfield and Anfinsen (1956) limited the specificity still further. They were engaged in the study of ribonuclease, which had previously been oxidized with performic acid to break the disulfide bonds, and chose to treat the protein with 2,4-dinitrofluorobenzene, thus forming the ϵ -DNP derivatives of all the lysine residues present in the molecule (and of course the α, ϵ -di-DNP derivative of the lysine residue at the N-terminal end of the molecule). This renders the lysine residues resistant to the action of trypsin, leaving the arginyl residues as the only point of attack for the enzyme. This procedure of course gives a relatively very small number of peptides (five in the case of ribonuclease) with an *average* chain length of the order of 20 or 25 residues each. If these are then fractionated, each of the separate products may then be further analyzed and fractionated by such methods as those employed by Sanger for the chains of insulin.

Chymotrypsin has an entirely different specificity of attack than trypsin, splitting primarily peptide bonds involving phenylalanine or tyrosine residues, although occasionally other points are attacked. Pepsin, papain, and a proteolytic enzyme from *Bacillus subtilis* (subtilisin) have also often been used in these studies, and other enzymes as well. We attempt no discussion of details here; the situation is well reviewed by Anfinsen and Redfield (1956).

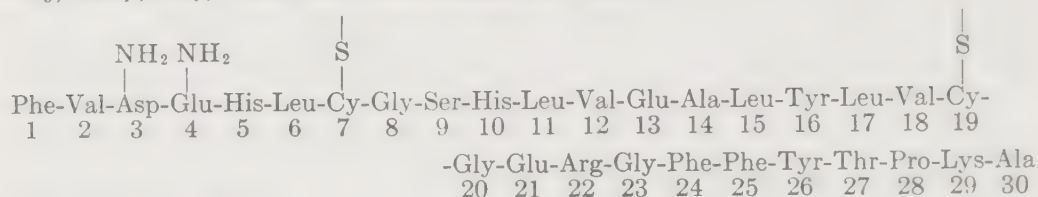
We shall not consider further the techniques involved nor the many difficulties and pitfalls that confront the investigator in this field. These must be learned from careful study of the work of the original investigators and, for those deeply concerned with these matters, by personal experience. Here we present some of the results hitherto obtained on naturally occurring peptide chains (Table IV).

All the peptide chains shown in Table IV represent hormones or (in the case of the insulins) chains derived from hormones by oxidation of disulfide bonds. No theory has yet been put forward to correlate hormone activity with structure. The sequences are set down here partly as the record of a great triumph of scientific achievement in unraveling structures which might once have appeared so complex as to be inaccessible to attack, but also as riddles, the true meaning of which to the biologist has not yet been read. The reader should look, and ponder.

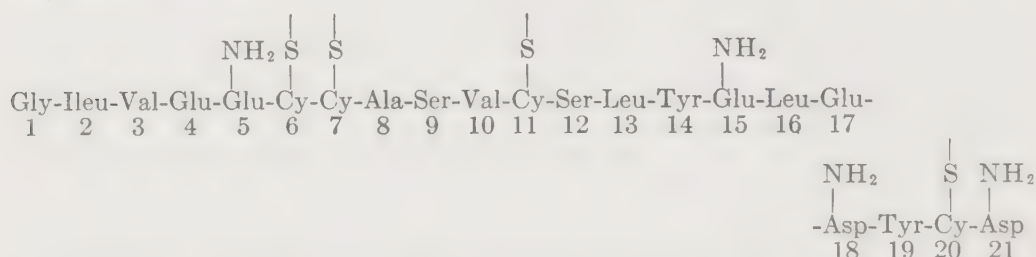
It should be said at once that the relatively short chains portrayed in Table IV are probably not typical of those to be found in the larger proteins. It will be noted, for instance, that except in the insulin chains none of the molecules represented contains either cystine or cysteine residues. In general, from these and all the other data available at present, it would appear that almost any residue may be followed by any other, and there seems to be no tendency for sequences of a given set of amino

TABLE IV
SOME POLYPEPTIDE CHAIN SEQUENCES IN HORMONES OF PITUITARY AND
PANCREAS

Hog, Sheep, Beef, Horse, and Whale Insulin, B Chain



Beef Insulin, A Chain



Comparison with Insulins of Other Species. The B chain of insulin is identical in beef, hog, sheep, horse, and whale. The A chain is also identical in all these species, except for residues 8, 9, and 10, within the intrachain disulfide loop. The sequence for these residues is as follows:

Species	Position		
	8	9	10
Beef	Ala	Ser	Val
Hog	Thr	Ser	Ileu
Sheep	Ala	Gly	Val
Horse	Thr	Gly	Ileu
Whale	Thr	Ser	Ileu

(Note that hog and whale insulins happen to be identical.)

Glucagon (Hyperglycemic-Glycogenolytic Hormone of Pancreas)

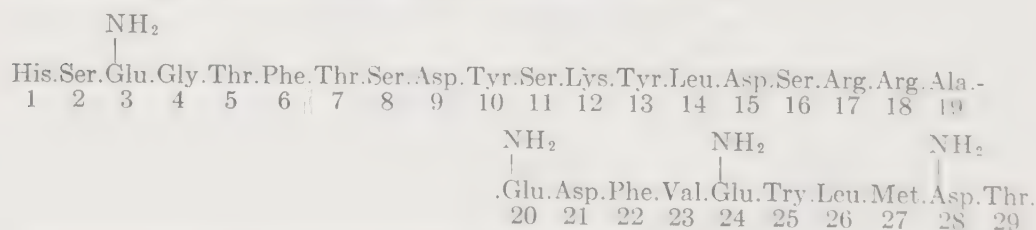
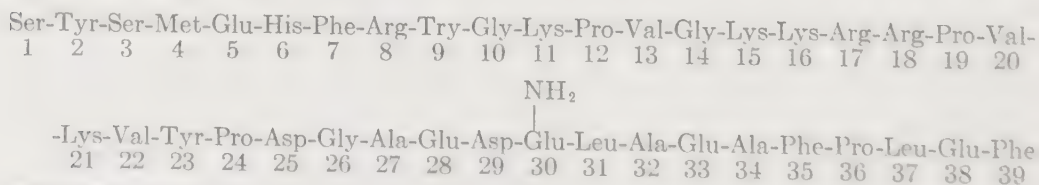
 β -Corticotropin (Hog)

TABLE IV (*Continued*)*α-Corticotropin (Sheep)*

Sequence same as for β -corticotropin with the following exceptions:

Ala-Gly-Glu-Asp and Ala-Ser.
25 26 27 28 31 32

Residue 30 may be glutamic acid in the sheep hormone, rather than glutamine, as reported for the hog hormone.

Hog Melanocyte-Stimulating Hormone (MSH, Melanotropin)

Asp.Glu.Gly.Pro.Tyr.Lys.Met.Glu.His.Phe.Arg.Try.Gly.Ser.Pro.Pro.Lys.Asp
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

Beef Melanocyte-Stimulating Hormone (MSH, Melanotropin)

Sequence same as in hog MSH, except that Glu in position 2 is replaced by Ser.

References: Beef, hog and sheep insulins: Brown *et al.* (1955); horse and whale insulins: Harris *et al.* (1956); glucagon: Bromer *et al.* (1957); β -corticotropin (hog): Bell (1954), Howard *et al.* (1955); α -corticotropin (sheep): Li *et al.* (1955); on corticotropins see also Li (1956); hog MSH: Harris and Roos (1956), Geschwind *et al.* (1957a); beef MSH: Geschwind *et al.* (1957b), also Li (1957).

acids to recur at regular intervals in any given chain. On the other hand, proteins with corresponding functions from different animal species have, as might be expected, almost identical sequences, to judge from Sanger's data on insulin. The sequence in the B chains is identical in the three species of insulin studied; the A chains differ only in the residues in positions 8, 9, and 10, which lie between the half-cystine residues in positions 6, 7, and 11. Far more knowledge is needed, of course, before generalizations concerning structural relations of similar proteins in different species can be made.

It is interesting to note that a sequence of seven residues—Met·Glu·His·Phe·Arg·Try·Gly—is found in positions 7 to 13 inclusive of the melanophore-stimulating hormone, and also in positions 4 to 10 inclusive of the corticotropins. Since the latter show some melanophore-stimulating activity, as well as their corticotropic activity, the relation is interesting (Harris and Roos, 1956). In general it is a striking feature of the structure of the corticotropins that the positively charged groups appear in the N-terminal half of the molecule, the negatively charged groups near the C-terminal end. Three hydroxyamino acid residues (Ser·Tyr·Ser) appear at the extreme N-terminal end, and any tampering with the integrity of the molecule at this end leads to complete loss of hormone activity. On the other hand, at least 10 or 11 residues can be removed by enzymatic action from the C-terminal end, apparently without any loss of hormone action. The interest of these findings is of course very great, but their significance is still obscure. The reader who wishes to learn

more of these hormones may begin by consulting the recent review of Li (1956).

Glucagon, which acts to raise the glucose content of the blood—in contradistinction to insulin, which lowers it—contains a strikingly large number of residues with charged side chains, or with side chains capable of acting as hydrogen bond donors. Of the 29 residues present, 21 belong in these two categories.

Numerous short sequences have been determined in the products of partial hydrolysis from lysozyme, collagen, silk fibroin, keratin and various other proteins. Many of these are tabulated, and references to others are given, by Anfinsen and Redfield (1956). We shall have occasion to refer to some of these data in discussing the spatial configurations of peptide chains in the fibrous proteins.

Disulfide Linkages in Insulin and Ribonuclease

The determination of the positions of the disulfide linkages in the insulin molecule proved a particularly difficult problem, finally solved by Ryle *et al.* (1955), after the structure of both the A and the B chains had been worked out in detail, as shown in Table IV. Four of the half-cystine residues are in the A chain, only 2 in the longer B chain. The molecular weight determinations (Harfenist and Craig, 1952) indicate that only one chain of each type is present, and the analytical evidence shows that all the half-cystine residues are joined in pairs to form disulfide linkages; insulin contains no free sulfhydryl groups. (Indeed reduction of any of the disulfides to sulfhydryl groups destroys the hormone activity.) Thus it may be presumed that the 2 CyS— residues at positions 7 and 19 of the B chain form disulfide cross-links with 2 of the 4 residues at positions 6, 7, 11, and 19 in the A chain. Then the other 2 residues of the A chain must form an intrachain disulfide link which makes a loop in the chain. Great difficulties were found in determining the sequences adjoining the S—S links in the original insulin molecule largely because of rearrangements that occur due to disulfide interchange reactions. Such reactions may be denoted by the general scheme



These interchanges can occur in neutral solutions, in which they are catalyzed by sulfhydryl compounds, and also in strongly acid solutions, in which they are actually inhibited by sulfhydryl compounds (Ryle and Sanger, 1955). After the conditions for inhibiting the interchange reactions were worked out, it was possible to determine a unique set of relations for the disulfide cross-links; the results are shown in Fig. 5. The intrachain disulfide link in the A chain joins residues 6 and 11; the loop

in the A chain so produced involves a 20-atom ring of the same type found in oxytocin and vasopressin (Fig. 4). We return later to the implications of this structure for the stereochemistry of insulin.

To determine the complete structure of the insulin molecule, it was necessary to determine the allocation of the six amide groups among the aspartic and glutamic acid residues. This was an undertaking of considerable difficulty, for all the amide linkages are broken on acid hydrolysis, and ammonia is liberated. The amide linkages remain intact during enzymatic hydrolysis, however, and by separation of the peptides produced by such hydrolysis, and determination of the content of amide groups in each, it was possible to work out a unique assignment of the six amide groups in insulin.⁶ This has already been shown in Table IV. It is seen from Table IV that the three aspartic acid groups, obtained on complete hydrolysis of the protein with acid, actually exist as asparagine residues

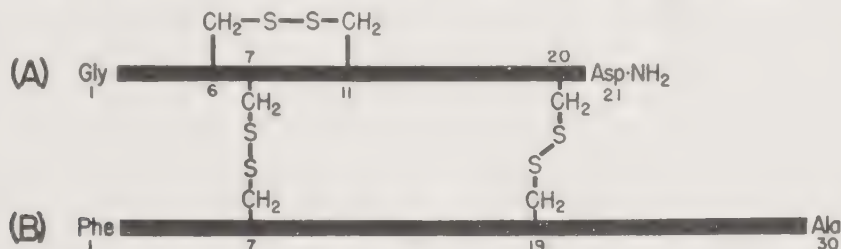


FIG. 5. Outline of the structure of beef, hog, and sheep insulins. The glyceryl (A) chain, with 21 residues, and the phenylalanyl (B) chain, with 30 residues, both have the same orientation, with the free terminal α -amino groups at the left hand side of the diagram. The complete sequence of the residues in the A and B chains is given in Table IV.

in the protein, with amide groups (not free carboxyls) at the end of the side chain. Of the seven glutamic acid groups obtained on complete hydrolysis, three were present as amide (glutamine) groups in the protein, the other four as glutamic acid residues with free carboxyls.

*Structure of Ribonuclease.** The molecule of bovine pancreatic ribonuclease—an active enzyme which digests ribonucleic acid with the formation of oligonucleotides—is more than twice as large as that of insulin. All the 124 amino acid residues, however, are linked in a single peptide chain, which is internally cross-linked by four disulfide bonds. A series of outstanding researches (among which we may name Hirs, *et al.*, 1960; Spackman *et al.*, 1960; Smyth *et al.*, 1962; Potts *et al.*, 1962; Gross and Witkop, 1962) have led to a complete formulation of the chain sequence, and to the location of the disulfide cross-links. The sequence studies were carried out after the disulfide groups had been oxidized, with breakage of

* Actually other procedures were also used in working out the complete assignment (Sanger, 1956), but we shall not discuss them here.

* The discussion on pp. 97–99 has been revised in the second printing of this book (1962).

the crosslinks; the resultant oxidized ribonuclease, which was enzymatically inactive, could then be readily digested by proteolytic enzymes such as trypsin and chymotrypsin, to yield a number of well defined peptides. These peptides could be separated by column chromatography on

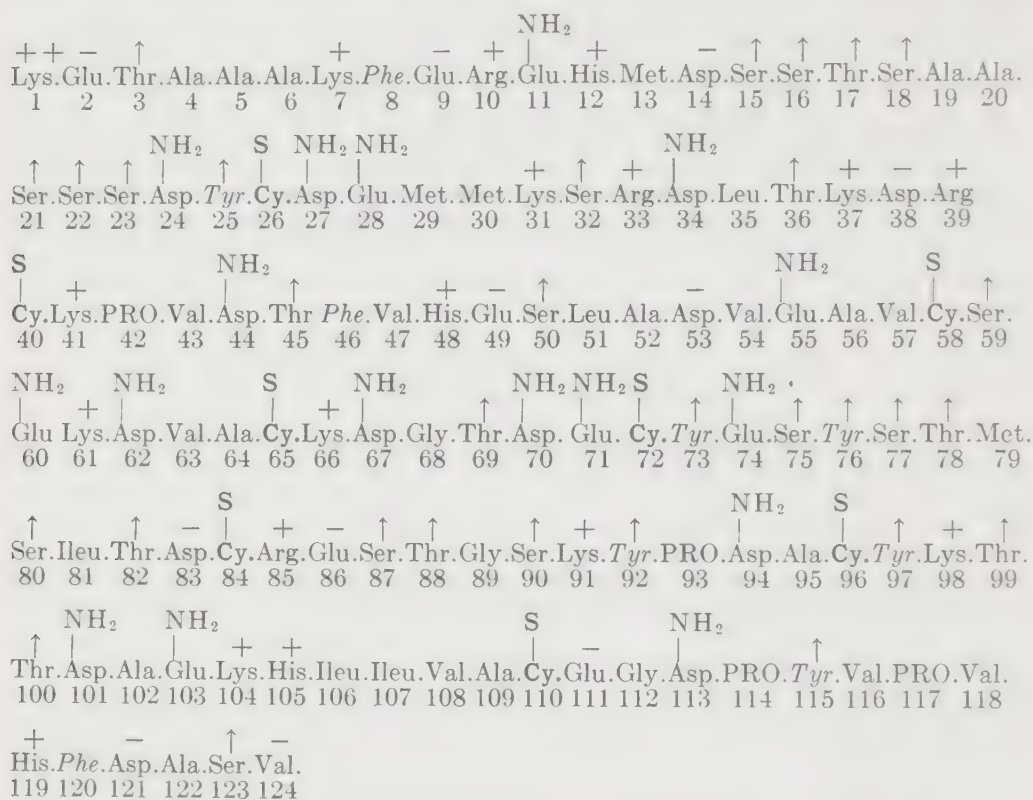


FIG. 6. The amino acid sequence in the peptide chain of ribonuclease. Hydroxyl side chains are denoted by a vertical arrow, amide groups by $-NH_2$ above Asp or Glu, positively and negatively charged groups by + and - respectively. Aromatic side chains (*Phe* and *Tyr*) are italicized, prolyl residues are denoted by PRO. The four disulfide bonds are formed by linkage between the 8 *CyS*-residues, printed in bold face type. The links join residues 26 and 84, 40 and 96, 58 and 110, 65 and 72.

ion exchange resins or other supporting media. The individual peptides were studied, and their sequences determined, by methods similar to those already described; the Edman degradation (p. 83) proved extremely valuable in sequence determination. Determination of sequential relations between the different peptides was achieved by successive digestion with different enzymes, applied in different order; this permitted determination of overlapping sequences. The determination of the specific cross-links, joining the 8 half cystine residues to form 4 disulfide bonds, presented great difficulties, as with insulin (p. 96) but was finally achieved (Spackman *et al.*, 1960). All these researches have led to the proposal of a

detailed structure (Fig. 6), which may be subject to slight revision in the future, but is almost certainly very close indeed to the correct formula.

The enzyme subtilisin, under carefully controlled conditions, attacks a single peptide linkage in ribonuclease, between residues 20 and 21, thus yielding a peptide (S-peptide) of 20 residues and a protein (S-protein) of 104 residues. Either component is inactive when separated from the other, but when mixed the two adhere to one another with a very high association constant, to form an active enzyme once more. Richards and Vithayathil (1959, 1960, 1961) who discovered these important facts, have investigated numerous effects on chemical modification of S-peptide and S-protein, with respect to the affinity of the components for one another, and the activity of the resulting complex (see also Allende and Richards, 1962).

The disulfide bonds of ribonuclease may be broken by reduction with mercaptoethanol in urea solution. The reduced enzyme is completely inactive, and appears to be unfolded into the form of a random coil. On removal of the urea, and with slow reoxidation in air, the disulfide bonds rejoin in the correct arrangement and enzyme activity is fully restored (White, 1961; Haber and Anfinsen, 1962). This indicates that the nature of the primary sequence of amino acid residues, in neutral aqueous solution determines the three-dimensional folding of the molecule into its native conformation. A thoughtful discussion of the whole situation has been given by Anfinsen (1962).

Most protein molecules are larger—many are vastly larger—than insulin or ribonuclease, and it is highly likely that many different types of design are employed in their construction. For instance, the hemoglobins, which exist as stable molecules with 4 peptide chains and 4 heme groups in neutral aqueous solution, break up readily into subunits in acid, or in 6 *M* urea. Clearly the subunits are not held together by covalent bonds; and indeed there are no disulfide bonds in human or horse hemoglobin. Detailed sequence studies of the two kinds of peptide chains (α and β chains) in normal adult human hemoglobin have now been carried out (Braunitzer *et al.*, 1961; Konigsberg *et al.*, 1961; Goldstein *et al.*, 1961). These will be discussed in Volume II, in their relation to the three-dimensional conformation of hemoglobin.†

Spatial Configurations of Polypeptide Chains

The studies just described have led to complete structural formulas for the corticotropins, glucagon, and three species of insulin, to something approaching such a formula for ribonuclease, and to much important although incomplete information regarding the peptide chain sequences

† For the references to recent work (1959–1962) on ribonuclease and hemoglobin, see p. 662.

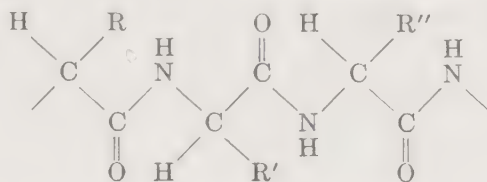
in other proteins. They do not, however, give direct information concerning the stereochemistry of the peptide chains. To obtain a complete picture of the structure of such a molecule, we should be able to specify the coordinates of the atoms in space, which involves a knowledge of the interatomic distances and bond angles, and of the extent to which something like free rotation is permissible about the various bonds. Such knowledge is important for the understanding of the biological functioning of these molecules. For instance, all enzymes are proteins, and there is good evidence that the combination of an enzyme with its substrate generally involves at least three different points of attachment on the enzyme surface, which must come into contact with three complementary portions of the substrate molecule. The exact spatial arrangement in the neighborhood of the active center of the enzyme is therefore of crucial importance. In the interaction of antigens, which are generally proteins, with antibodies, which are always proteins, a similarly precise fitting together of complementary portions of antigen and antibody is fundamental.

As yet we do not know enough to answer such questions, which are of burning interest to the biologist, except in rather vague terms. Detailed knowledge is available, however, concerning the dimensions of polypeptide chains and some of the coiled or extended forms which they can assume. This knowledge is derived chiefly from X-ray diffraction studies on crystalline amino acids and small peptides, and on fibrous and crystalline proteins. To present the nature of the X-ray evidence with anything approaching adequacy would require writing not only a chapter but a large portion of a book in addition to this one. This task will not be attempted here. We call the attention of the reader to some of the many treatments of the subject which are available elsewhere, and urge him to become familiar with at least the basic elements of the subject, preferably to study it more deeply. It seems almost certain that X-ray diffraction will contribute more than any other physical tool in unraveling the structure of complex, highly organized macromolecules, such as the proteins and nucleic acids. Here, however, we shall simply assume the basic information regarding interatomic distances and bond angles, as we have done already in Chapter 2 in discussing the structure of water.

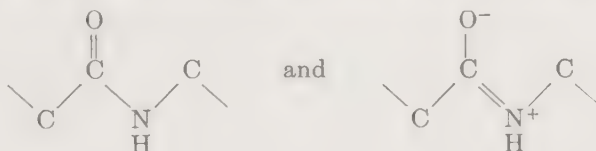
For the reader who wishes to pursue the subject of X-ray diffraction, the books by Bragg (1931), Bunn (1945), Lonsdale (1949), and Robertson (1953) may be recommended. Bragg's book is probably the most general in its scope; Bunn's approach is from the point of view of the chemist, and many chemists may find the subject easier to grasp as presented by Bunn; Lonsdale's book is a compact and, on the whole, lucid presentation of the fundamentals of X-ray crystallography; Robertson's book consists of two parts, the first dealing with general crystallographic principles, the second giving a very valuable account of the structure of many important organic molecules. A brief and interesting introduction to the X-ray study of biological

macromolecules is given by Kendrew and Perutz (1949), but the subject has made rapid strides since this article was written, both with respect to obtaining the experimental data and with respect to the speed and power of the computing techniques needed to evaluate the data. The results of X-ray studies on amino acids, peptides, and proteins are well presented by Low (1953), who also provides much of the fundamental background for understanding the methods involved. Recent work on proteins has been reviewed by Kendrew (1954a) and by Crick and Kendrew (1957). See also the Royal Society discussion on proteins (Astbury and others, 1953), in which the helical configurations of peptide chains, proposed by Pauling and Corey, are thoroughly discussed; some of the points which were then a matter for debate have since been settled.

The essential unit of the polypeptide chain is the repeating sequence:



The C—C bonds of the chain are found, as expected, to be normal single bonds with a C—C distance of approximately 1.53 Å. Likewise the C—N bonds between the CHR and NH groups are single bonds, with a distance close to 1.47 Å. Freedom of rotation around these bonds should be comparable to that found in ordinary hydrocarbon chains. The C—N bonds in the —C(O)—NH— groups, however, have a large amount of double-bond character, due to resonance between the structures:

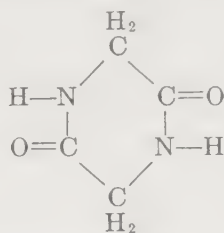


This is confirmed by the studies on simple peptides, which show that the C—N distance is near 1.32 Å, far shorter than the normal single-bond distance. This partial double-bond character of the C—N linkage has two important effects: (1) The entire group of atoms in the formula above is constrained to assume a coplanar, or nearly coplanar, configuration, as in a structure like ethylene. These stable configurations may be either *cis* or *trans*:



In general the *trans* configuration appears to be favored; it is the one

found in all the peptides for which exact structure determinations are available. The *cis* configuration is found in diketopiperazine:



(2) The distribution of electric charge is such that there is an excess of negative charge on the oxygen of the $C=O$ group, beyond that characteristic of simple carbonyl compounds (aldehydes and ketones), and an excess of positive charge, and therefore a deficiency of electron density, in the $N-H$ bond. Hence the former has a very strong attraction for hydrogen-bonding donor groups, and the latter for acceptor (nucleophilic) groups. Hydrogen bonds formed by the groups in the peptide linkage are therefore in general exceptionally strong. The attractions involved are best satisfied if each $C=O$ group can form a hydrogen bond with an $N-H$ group of another amino acid residue of the same, or of another neighboring, peptide chain; and likewise each $N-H$ group tends to form a similar linkage with a $C=O$ group of another residue.

The dimensions of the fully extended polypeptide chain have been estimated by Pauling and Corey and are shown in Fig. 7. It should be noted that, if all the residues in the chain have the same configuration around the α -carbon atom—that is, for peptide chains such as are found in natural proteins, the *L*-configuration—then the *R* groups in such a fully extended chain must point alternately upward and downward, at an angle of $110^\circ/2$, from the plane of the drawing. The repeat distance from a given *R* group to the next one pointing in the same direction is 7.23 Å; this distance of course corresponds to two residues along the chain, or 3.62 Å per residue. This configuration, however, is one which can seldom be realized in practice, for when a space model is constructed it is found that the side chains interfere with one another if the chain is fully extended in this fashion. Only in polyglycine, in which the *R* group on the α -carbon atom is a hydrogen atom, is this interference absent. The configurations which we may expect to be realized in actual substances must be less fully extended.

In seeking for possible stable structures, we look first for patterns involving regular repetitions of the configurations of the peptide residues along the chain. That is, if a given operation, usually the combination of a translation and a rotation, carries residue number 1 into the position of residue 2, repetition of the same operation will carry it into the position

of residue 3, and so on indefinitely. We impose certain restrictions which any plausible configuration must obey: (1) The covalent interatomic distances and bond angles shown in Fig. 7 must be maintained within narrow tolerances, say within ± 0.01 Å for the distances, and ± 3 or 4

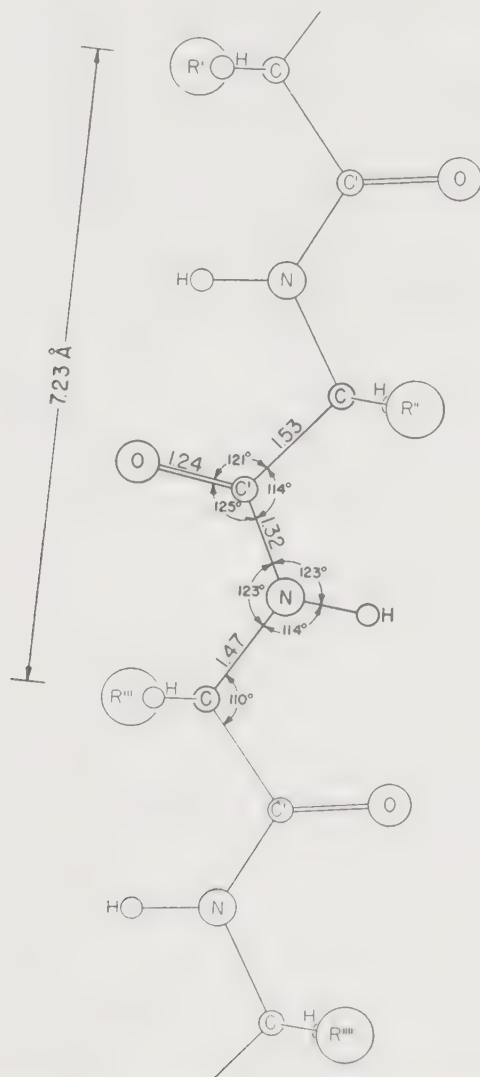


FIG. 7. Dimensions and bond angles of a fully extended *trans* polypeptide chain. (From R. B. Corey and L. Pauling, *Proc. Roy. Soc. B* **141**, 10 (1953).)

degrees for the angles.⁷ (2) The atoms involved in the amide groups must remain coplanar, or nearly so, for reasons discussed above. (3) For maximum stability, each N—H group must be hydrogen-bonded to a C=O, and each C=O to a N—H. We note that this condition cannot be

⁷ As a rule, in constructing models the distances are considered fixed, but a slight distortion of the angles is permitted.

30% protein; it is not, however, a true solution, for it can be stretched, and behaves at first like rubber, showing reversible long-range elasticity. On repeated stretching, or on being held under tension, however, the fibers lose their elasticity and acquire the mechanical properties of silk; the characteristic X-ray diffraction pattern appears also at the same time. This X-ray pattern has been studied repeatedly, beginning with the work of R. Brill in 1923. The history of the developments and the present state of our knowledge of the structure are well presented by Kendrew (1954b) and by Marsh *et al.* (1955). The main points with which we are concerned here are that the diffraction pattern indicates a high degree of organized crystalline structure in the fiber, and that there is a repeating spacing, 7 Å in length, along the fiber axis. Moreover, the silk fiber, though somewhat extensible, cannot be stretched very far without breakage or slipping within the fiber. This suggests—and a detailed analysis of the X-ray pattern and other data justifies the suggestion—that the polypeptide chains of silk are in a configuration which is not far from the fully extended state portrayed in Fig. 7. The repeating distance for two amino acid residues is a little less than in that figure, 7.0 instead of 7.23 Å, but the difference is small. Furthermore, silk is mostly made up of amino acid residues with very small side chains: glycine, alanine, and serine. These can pack quite closely into superimposed sheets of linked polypeptide chains, although the presence of some larger residues, especially tyrosine, with its bulky phenyl groups, complicates the situation. It cannot be claimed that a final solution to the problem has been attained, but a structure which seems to meet nearly all the requirements has been proposed by Marsh *et al.* (1955). This is based on a pattern of parallel extended polypeptide chains, cross-linked by $\text{—C=O} \cdots \text{H}\cdot\text{N—}$ hydrogen bonds between adjacent chains. The direction of the hydrogen bonds is very nearly perpendicular to the direction of the axis of the chains. The structure, which was worked out earlier by Pauling and Corey (1951, 1953) from general considerations such as we have outlined above, is known as an antiparallel pleated sheet and is shown in Fig. 9.

As may be seen from the drawing, any one peptide chain in this structure has a configuration not far from that of the fully extended chain in Fig. 7. The atoms in the main peptide chain are not completely coplanar, however, as they are in the fully extended structure. Figure 9 shows that the β -carbon atoms of the R groups, which are attached to the α -carbon atoms in corresponding positions in adjoining chains, face alternately toward one another and away from one another. In either case they do not point in the same direction, as they do in another closely related structure—the parallel chain pleated sheet which is described below. The repeating distance for two peptide units along the chain in

the antiparallel pleated sheet is 7.0 Å, which corresponds almost exactly with the observed distance (6.97 Å, according to Marsh *et al.*). The N—H · · · O hydrogen bonds are practically linear and are 2.76 Å in length. The distance between the axes of adjoining peptide chains in the plane of the sheet is 4.7 Å, and the R groups project above and below this plane. The packing of the chains, in the direction perpendicular to this plane, is therefore dependent on the size of the side chains. Small side chains, like those of glycine, alanine, and serine, the principal constituents of silk fibroin, can pack more closely than larger ones. Several

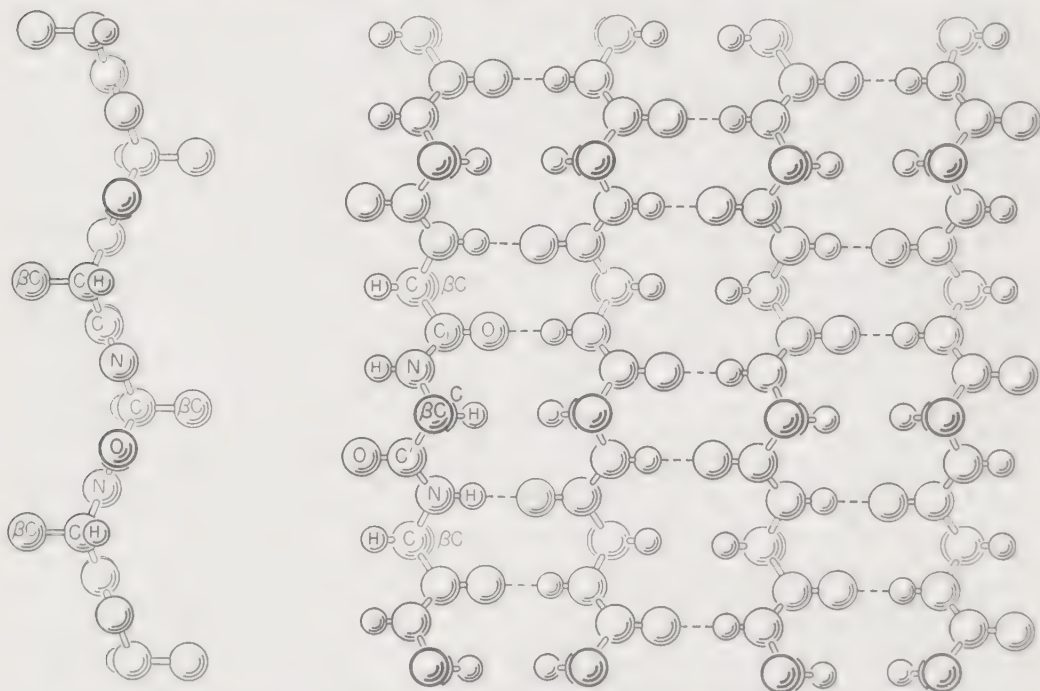


FIG. 9. The antiparallel chain pleated sheet structure of polypeptide chains. (From Marsh, Corey and Pauling (1955).)

drawings of possible packing models for silk fibroin are given by Marsh *et al.*, who have also given a detailed critical analysis of the observed and calculated X-ray diffraction patterns. We shall not discuss this in detail here. The antiparallel chain pleated sheet, however, does appear to provide an excellent basis for describing the structure of silk fibroin.

β -Keratin

Keratin, the fibrous protein of horn, wool, hair, and related structures, is quite different in its properties. The amino acid analysis (Table II) is notable for the high proportion of cystine disulfide linkages present. Nearly all the known amino acids are present in significant amounts, although there is very little histidine, tryptophan, or methionine. Unlike

silk fibers, wool and hair are highly extensible materials. They can be stretched to twice their original length or more under suitable conditions, and may shorten again to their original length on removal of the tension. In some classical studies, which are still well worth reading today, Astbury and his collaborators (1931, 1933) showed the relation between the changes in the X-ray diffraction pattern and the mechanical changes produced by reversible stretching. Keratin in its natural unstretched state is called α -keratin and exhibits a characteristic X-ray diffraction pattern, known as the α -pattern. On stretching, it is converted into what is known as β -keratin and exhibits quite another X-ray pattern, the β -pattern. Complete stretching of hair or wool to twice the original length is generally achieved only in the presence of moisture at temperatures near 100°. The combined action of moisture and heat may be due largely to rupture of the disulfide cross-links, which must probably be broken before the peptide chains of keratin can fully unfold. The phenomena are complex and will not be discussed here; a good survey of the status of the problem is given by Kendrew (1954b).

What concerns us now is the structure of the extended β -keratin. Here the X-ray measurements show a repeating spacing along the fiber axis of 3.33 Å. Astbury concluded, and all later workers have agreed, that this represents the distance corresponding to a single amino acid residue in the direction of the chain axis. There are two other well-marked X-ray reflections at right angles to this axis, of 9.7 Å and 4.65 Å, respectively. Moreover, in keratin fibers which have been not only stretched into the β -form but also pressed and rolled so as to produce alignment of the chains relative to one another, it is found that the 9.7-Å and 4.65-Å spacings are at right angles to one another as well as to the 3.33-Å spacing. Again the general interpretation seems fairly clear. The spacing at 4.65 Å is virtually identical with the distance of 4.7 Å between the axes of chains in the same sheet, found in silk fibroin—what Astbury has called the backbone spacing. The perpendicular distance of 9.7 Å must then represent the distance between adjacent parallel layers of peptide chains. It is determined largely by the distance to which the side chains project outward from the main peptide chains.

The side-chain spacing of 9.7 Å is the value for dry keratin; an increase of several per cent is observed when the keratin is soaked in water. Presumably the water penetrates between and among the side chains; the charged and polar groups of the side chains bind water strongly, and the structure expands somewhat in this direction to accommodate the water molecules. A diagrammatic picture of the β -keratin structure, as envisaged by Astbury, is given in Fig. 10.

Perhaps the most plausible structure for β -keratin, in the light of our

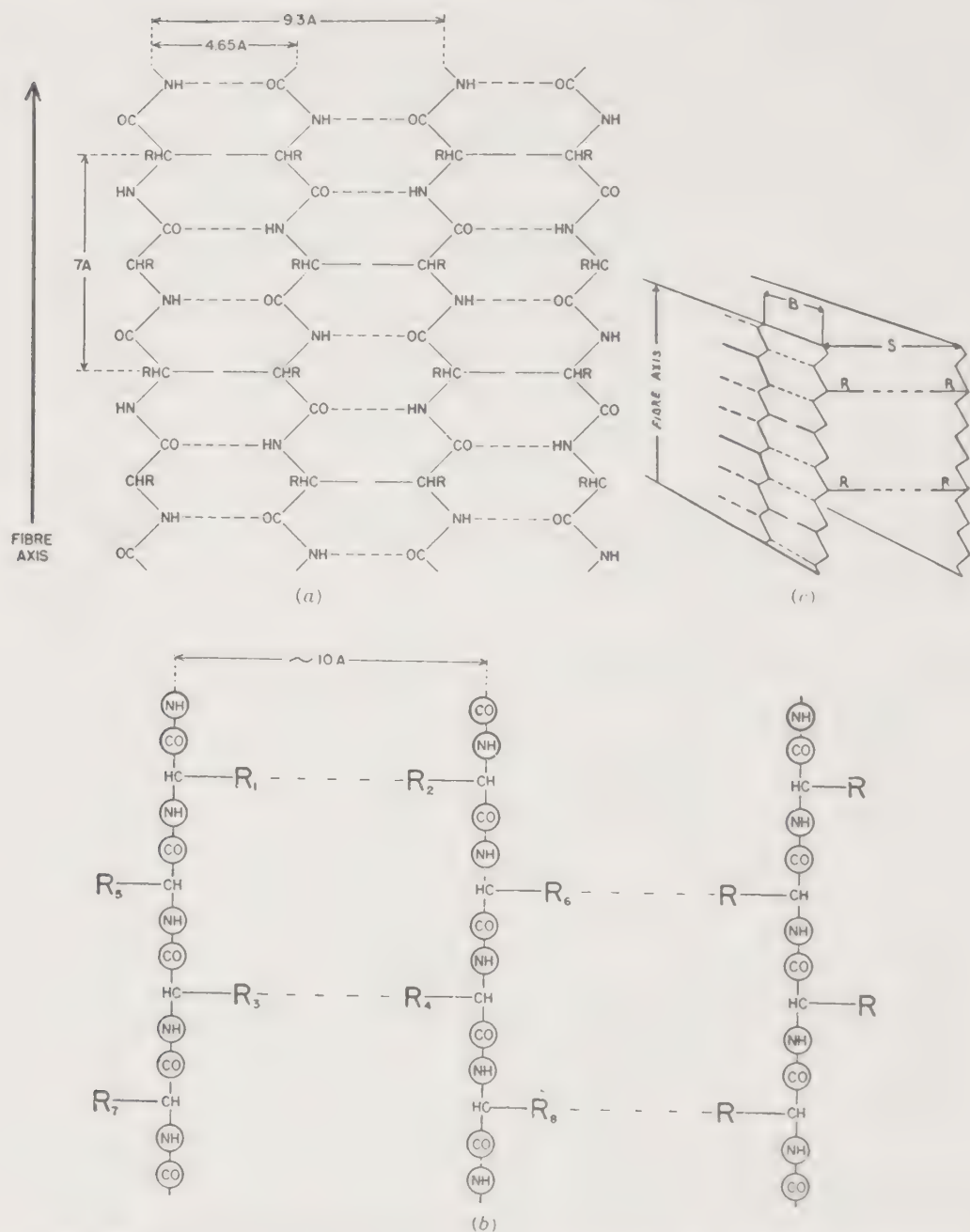


FIG. 10. General features of the structure of β -keratin. (a) View of a single layer of close packed chains showing backbone spacing (4.65 Å). (b) View parallel to the grid formed by polypeptide layers showing the side-chain spacing. (c) Key showing relation between backbone layers and side-chain grid perpendicular to them. (From B. W. Low (1953), p. 257; after drawings in W. T. Astbury and H. J. Woods (1933).)

present knowledge, is another of the structures proposed by Pauling and Corey (1951, 1953), the parallel chain pleated sheet. This is shown in Fig. 11. Its similarity to the antiparallel chain pleated sheet, in so far as the arrangement of the peptide chains is concerned, is immediately apparent. The differences between the relations of corresponding groups in adjacent chains are also apparent. This picture is essentially in harmony with the earlier model proposed by Astbury and illustrated in Fig. 10; but it goes further by proposing precise positions for all the atomic coordinates in the repeating unit of structure. Given such precise coordinates, X-ray diffraction patterns can be calculated, and the intensity of

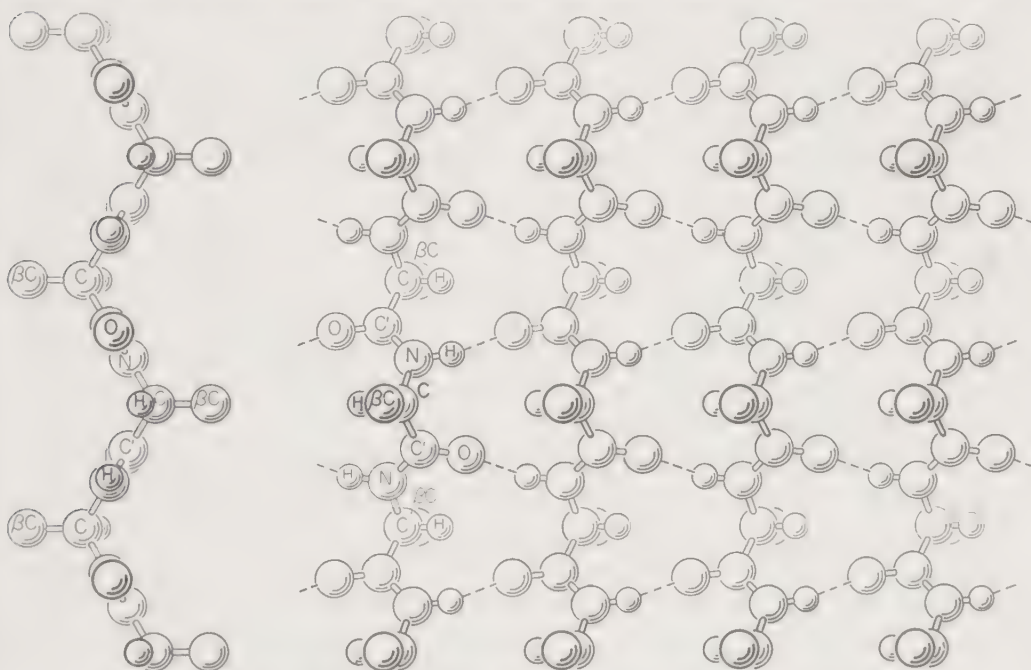


FIG. 11. The parallel chain pleated sheet structure of Pauling and Corey.

the observed reflections compared with the calculations. This is essentially what Marsh *et al.* (1955) have done for silk fibroin, as we have discussed above, with highly encouraging results. In β -keratin the problem is more difficult for two reasons: (1) The side chains are far more various than in silk fibroin; almost every kind of amino acid residue is present in significant amounts; and only a fragmentary knowledge of the sequence of residues in keratin is available. (2) There are large amorphous regions in keratin, and the X-ray diffraction patterns are therefore produced by only a portion of the structure, which is well ordered and crystalline. The orientation of the chains, even in the crystalline portion of the fibers, is less perfect than in silk fibroin, and the diffuse X-ray scattering from the amorphous regions blurs the pattern still more. Hence the observed X-ray

reflections are fewer and less sharply defined than in silk fibroin, and there is less information available for comparing a calculated with an observed structure. Therefore it certainly cannot be said that the problem of the structure of β -keratin has been solved in detail; but the working model of a system of parallel peptide chains, arranged in the pleated sheet formation, appears to be close to the truth.

α -Keratin and Synthetic Polypeptides: the α -Helix and Other Possible Helical Structures

The native form of keratin, α -keratin, is obviously made up of peptide chains which are somehow folded or coiled into a more compact configuration than the extended chains of β -keratin. The nature of the coiling, however, has been a matter of debate for many years; the history of the problem is well described by Low (1953) and by Kendrew (1954b). As we have already indicated in discussing the general possible configurations of peptide chains, a coiled structure, in contrast to an extended structure, is likely to be held in its coiled configuration by hydrogen bonds between a $\text{C}=\text{O}$ group of one amino acid residue and an $\text{N}-\text{H}$ group of another residue *in the same peptide chain*. Such helical structures had been suggested about 1940, and later, by H. S. Taylor and by M. L. Huggins. The first model of this sort, however, which was constructed in accord with the rigorous requirements specified above, with respect to interatomic distances and bond angles, was described by Pauling and Corey in 1951. It is known as the α -helix and is portrayed in Fig. 12. Each NH group is hydrogen-bonded to a $\text{C}=\text{O}$ group 3 residues beyond it in the chain, and each $\text{C}=\text{O}$ to an $\text{N}-\text{H}$ 3 residues away. The directions of the hydrogen bonds are nearly, though not quite, parallel to the axis of the helix. The dimensions of the helix may be slightly modified by assuming small and reasonable variations in the $\text{N}-\text{H} \cdots \text{O}$ distance; but a distance between 2.7 and 2.9 Å is considered most likely. The distance along the axis of the helix per amino acid residue is 1.47 to 1.53 Å, varying with the assumed $\text{N}-\text{H} \cdots \text{O}$ bond distance, and the helix makes a complete turn for each 3.6 to 3.7 residues. Thus the pitch of the helix—that is, the vertical distance along the axis from any point on the helix to a corresponding point directly above it—is approximately 5.4 Å ($= 3.6 \times 1.5$). If the number of residues per turn is 3.6, then five complete turns will correspond to 18 residues, and the whole pattern will repeat at this interval, which corresponds to a distance along the axis of approximately 27 Å (i.e., 18×1.5). Such a repeating unit has been found in fibers made up from certain synthetic polyamino acids, such as poly- γ -methyl-L-glutamate; but small variations in the number of residues per turn, due for instance to variation in the amount of stress on the fiber while it is

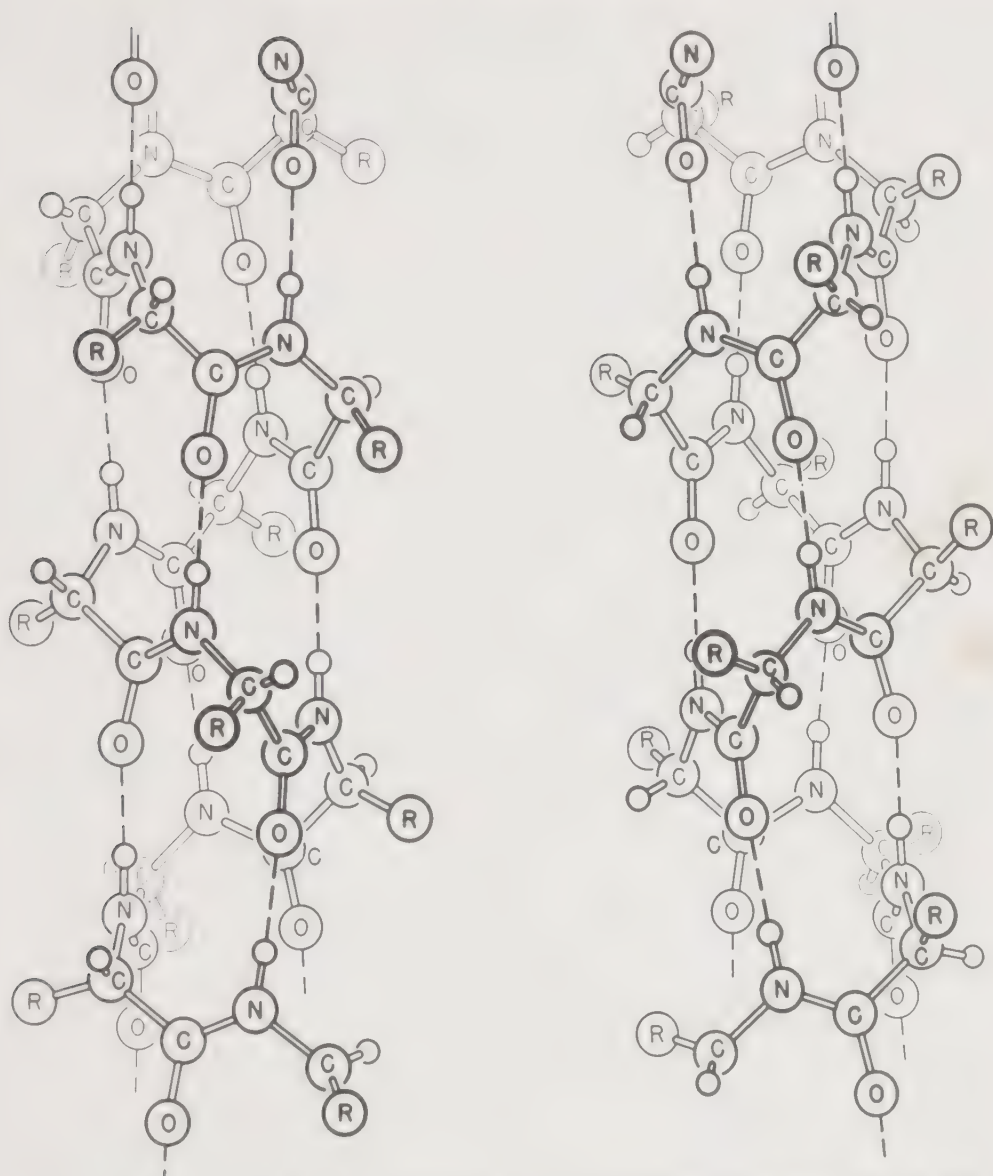
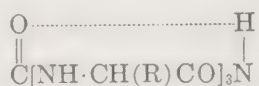


FIG. 12. Drawings of the left-handed and right-handed α -helices. The R and H groups on the α -carbon atoms are in the position corresponding to the known configuration of L-amino acids in proteins. (From a drawing provided by L. Pauling and R. B. Corey; see B. W. Low and J. T. Edsall (1956), p. 398.)

being prepared, may alter this value so that the repeating distance becomes much longer. This, however, although a matter of some crystallographic interest, is not important from the point of view of the structure of the helix.

The hydrogen bonds in the α -helix may each be regarded as closing a ring, in which 3 amino acid residues intervene between the C=O and

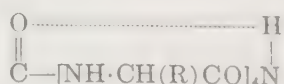
the NH group that are involved in the hydrogen bond. The sequence may be written



The reader should examine the sequences portrayed in the diagrams of Fig. 12 to verify this formula.

It is apparent from Fig. 12 that there are two ways of coiling an α -helix, corresponding to the turns of a left-handed and a right-handed screw. If all the R groups of the side chains were hydrogen atoms—that is, if we were dealing with a polyglycine helix—then it is readily apparent from the figure that the left- and right-handed helices would be mirror images of one another. As represented in the figure, however, all the α -carbon atoms in both helices are in the L-configuration; and the positions of the R groups and hydrogens on the α -carbons are *not* mirror images in the left- and right-handed helices. (A left-handed helix of L-amino acid residues *would* be the mirror image of a right-handed helix of D-amino acid residues.) Hence, if such a helix is formed entirely from L- (or entirely from D-) amino acid residues, the left- and right-handed helices are not enantiomers. Thus the work of forming a helix from an extended polypeptide chain will in general be different for a left- and for a right-handed helix; and we may expect that the two forms will not be equally stable.

The α -helix is only one of a number of possible helices which can be formed from polypeptide chains, in conformity with the general requirements already outlined concerning interatomic distances and bond angles. It is, however, the most stable and almost certainly the most important, in terms of its occurrence in nature. The possible helical configurations are discussed by Low and Grenville-Wells (1953); we reproduce from their work an illustration of one of the other proposed structures, called by them the π -helix (Fig. 13). This closely resembles the α -helix in general arrangement, but 4 amino acid residues, instead of 3, intervene between each CO and NH group which are linked by a hydrogen bond. Thus in Fig. 13 the CO group of residue 1 is hydrogen-bonded to the NH group of residue 6, the CO group of residue 2 to the NH group of residue 7, and so forth. The formula for a hydrogen-bonded loop in the π -helix is thus



The π -helix, like the α -helix, may exist in either a right-handed or a left-handed form. The helix shown in Fig. 13 is right-handed. The π -helix has approximately 4.4 amino acid residues per turn, and the distance per

residue along the helical axis is close to 1.14 Å. It is not so tightly coiled as the α -helix; there is a small hole running down the middle, which is too small even for the insertion of a water molecule. This tends to make the π -helix less stable than the α -helix, since the van der Waals attractions

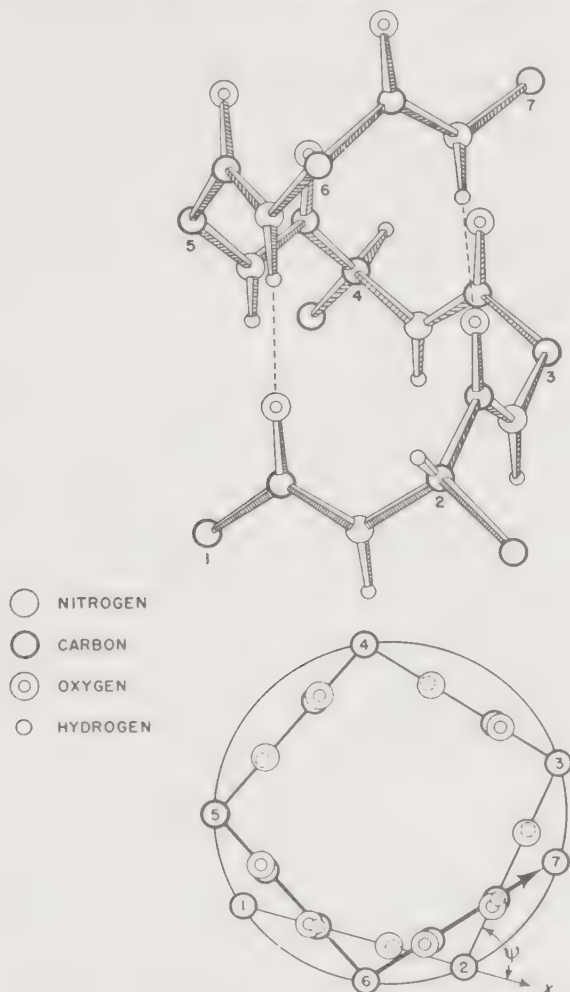


FIG. 13. The π -helix of Low and Baybutt (right-handed helix). Upper diagram, perspective drawing with the link-end carbon atoms numbered consecutively. The β -carbon and hydrogen atom positions of L-amino acid residues are shown attached to $C_2(\alpha)$ and $C_4(\alpha)$. Lower diagram, plan view of π -helix, looking down along the axis of the helix. The hydrogen positions are shown with broken lines. Nitrogen positions are omitted at the two crossover points. (From B. W. Low and H. J. Grenville-Wells (1953).)

between the atoms are better satisfied if they pack closely together on the inside of the coil.

The reader should note certain features common to both types of helix, which are perhaps illustrated most clearly in Fig. 13. In this figure, the

α -carbon numbered 1, the succeeding CO and NH group, and the α -carbon numbered 2 all lie in the same plane, according to the principle stated earlier (p. 101). The projection of this plane, looking down along the axis of the helix, is represented by the line from 1 to 2 in the lower half of the figure. Likewise α -carbon 2, α -carbon 3, and the intervening CO and NH group are all coplanar; the projection of this plane is given by the line from 2 to 3 in the diagram at the bottom.

A number of other possible helical configurations of peptide chains have been described; systematic exploration of the possibilities has been carried out by Donohue (1953) and by Low and Grenville-Wells (1953). For various reasons none of them seems at present particularly likely as a constituent of actual chemical structures. Without denying the possibility of their existence, therefore, we shall make no attempt to discuss them here.

Strong evidence that the α -helix exists as a constituent of synthetic or natural compounds was obtained from the X-ray study of α -keratin and also of fibers of certain synthetic polyamino acids, such as poly- γ -methyl-L-glutamate. The most crucial single piece of evidence is that these substances give a strong X-ray reflection corresponding to diffraction from planes normal to the fiber axis, and with a spacing of 1.5 Å. This corresponds to the distance along the axis per amino acid residue in the helix. The existence of this reflection was discovered experimentally only after investigators had been led to look for it because it was predicted as a consequence of the α -helix structure. None of the structures which had been proposed earlier for α -keratin predicted such a reflection; yet when found it was strong, as it should have been for an α -helix. In polymethyl-L-glutamate and other synthetic polypeptide fibers, a strong reflection corresponding to a spacing of 5.4 Å—representing the pitch of the helix—was also found. Moreover, in these synthetic materials, which are relatively simple in pattern since all the side-chain R groups are alike, the form of the entire X-ray diffraction pattern agreed in remarkable detail with the predictions for such a helical structure.

In α -keratin the agreement was less clear-cut, in spite of the striking appearance of the 1.5-Å spacing, for which no alternative explanation, other than the α -helix structure, has been found, and in spite of the general similarity of the X-ray pattern to that expected for a helical structure. The most serious difficulty was that the spacing of 5.4 Å did not appear; instead there was one near 5.1 Å, in a different position on the X-ray photograph from that predicted for the 5.4-Å spacing. Apparently, however, these discrepancies can be reconciled if it is supposed that the α -helices in α -keratin are coiled into bundles or cables of "coiled coils." Such proposals have been made independently by F. H. C. Crick and by Pauling and Corey (see Pauling, 1953, and the discussion by Low and

Edsall, 1956, pp. 404-406). Drawings of proposed ropes and cables, as given by Pauling and Corey, are reproduced in Fig. 14. We shall not present the arguments for and against such structures, since many technical intricacies are involved. They should certainly not be taken by the reader as proved, but they may well provide a basis for the interpretation of the structure of α -keratin. In any case, the evidence for the presence of the α -helix in the α -keratin structure appears virtually decisive.

It has been shown by Astbury that α -keratin is only one of a major family of proteins that are built on the same general plan. Myosin of

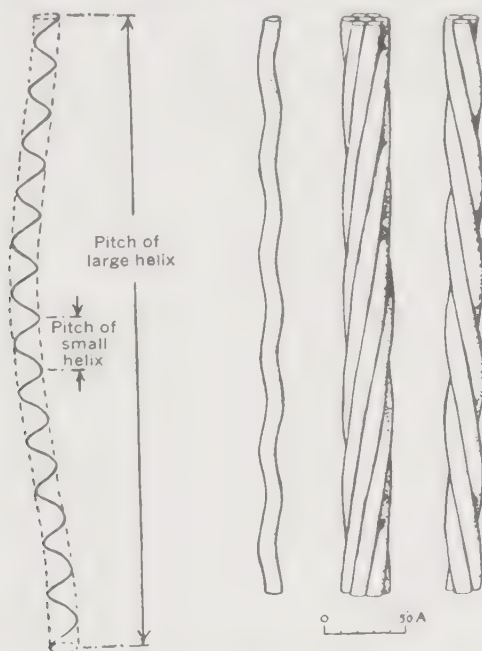


FIG. 14. Structure of compound α -helices, proposed to explain the structure of α -keratin. On the left, the coiling of the axis of an α -helix into a long helix. On the right, diagrams of a compound α -helix, of a 7 strand α -cable, and a 3-strand α -rope. (From Pauling (1953).)

muscle, epidermin of skin, and fibrinogen of blood plasma all can be shown to display essentially the same X-ray pattern as α -keratin, and are capable under suitable circumstances of being stretched to give the β -pattern. We are therefore dealing with a type of structure which is distributed throughout nature in a wide variety of fibrous, or potentially fibrous, proteins. The same kind of structure *may* also exist in many (possibly most) of the compact "globular" protein molecules which are soluble in water or salt solutions.

Collagen

The collagens represent a major group of fibrous proteins, entirely distinct in character from the keratin-myosin-epidermin-fibrinogen

group. The principal fibrous protein constituents of mammalian connective tissues, tendons, and bones have long been classed as collagens. Collagen fibers are only slightly extensible, and chemically they are relatively inert; metabolic tracer studies have shown that the turnover rate of collagen *in vivo* is very slow indeed. Collagen serves to maintain the structural framework of the tissues, as in bone and connective tissue, or as in tendons to transmit to other structures the tensions set up within the muscle fibers. Collagen is converted to gelatin on prolonged heating with water or alkali. It is characterized by an unusual amino acid composition, and by a distinctive X-ray diffraction pattern, the most prominent feature of which is a repeating spacing of 2.86 Å along the fiber axis. We return to these characteristic features in the discussion below. Observation of the X-ray pattern has revealed the presence of collagenous proteins in a great variety of animals, including most of the phyla of the animal kingdom. The collagen of the skin, tendon, and swim bladder of teleost fishes, commonly known as ichthyocol, has proved a particularly useful material for structural studies on collagen in recent years.

Collagen is of great practical importance. The chemistry of leather manufacture is primarily the chemistry of the collagens of mammalian skins, and of the action of tanning agents upon them. There are a number of valuable reviews of the chemistry of collagen; here we may mention those of Bear (1952), Kendrew (1954b), and Gustavson (1956). Some of the recent advances in our understanding of the structure of collagen, discussed below, appeared too recently to be included in any of these three reviews.

AMINO ACID COMPOSITION

The results of amino acid analyses on mammalian skin collagen are given in Table II. Several striking features of the analyses are immediately apparent. The predominant amino acid is glycine, which makes up approximately one-third of all the residues. About one residue in four is either proline or hydroxyproline, which are present in comparable amounts. This is particularly remarkable since hydroxyproline is practically unknown in other animal proteins, although it is occasionally found in plants.⁸ Hydroxylysine, found in small quantities in collagen—6 or 7 residues out of every thousand—is not known to be present in any other protein. Aromatic amino acids are nearly absent—tyrosine and phenylalanine make up less than 2% of all the residues, and there is no tryptophan. Cystine and cysteine are likewise lacking; thus there is nothing analogous to the

⁸ The protein elastin, also found in connective tissue, contains some hydroxyproline, although much less than collagen, as well as considerable amounts of proline. See for instance Kendrew (1954b), pp. 946-949.

network of disulfide cross-linkages found in keratin. There are a moderate number of positively and negatively charged side chains, and the isoelectric point of collagen is near neutrality. There are substantial numbers of nonpolar side chains—alanine, valine, leucine, and isoleucine.

HEAT SHRINKAGE

One of the most distinctive characteristics of collagen is the contraction of the fiber on heating. (See especially Bear, 1952, pp. 121–127, and Gustavson, 1956, Chapter 9.) This generally sets in quite suddenly, and over a very narrow temperature range, which for mammalian collagens is near 60° or a little above. The amount of shortening may be very large, the heat-treated shrunken fiber being often only one-third to one-quarter of the length of the original fiber. The physical characteristics of the shrunken material are drastically changed; it shows rubberlike elasticity and its tensile strength is much less than that of native collagen. The X-ray diffraction pattern largely disappears, indicating a high degree of disorientation of the fiber elements. The X-ray pattern is, however, partially restored by stretching the shrunken fiber. The effect of heat is probably due to a rupture of bonds which hold the chains together in an ordered arrangement. The transition temperature is lowered by treatment with alkali and raised by substances such as ethanol and methanol. In many respects the phenomena are strikingly like those of heat denaturation of soluble proteins. Boedtker and Doty (1956) have found that soluble fish collagen undergoes a dissociation into subunits on heating, essentially similar to the denaturation of many proteins, at a temperature very close to that at which the same collagen undergoes heat contraction in the fiber form.

T. Takahashi (as reported by Gustavson, 1956, p. 224) has pointed out an interesting correlation between shrinkage temperature (T_s) and hydroxyproline content for the collagens of various fish skins. Fish such as cod and halibut, which live in deep, cold ocean water, have collagens with T_s near to, or below 40°, and hydroxyproline content near 7 per cent. Fish living in warmer water, whether fresh or salt, such as pike, perch, carp, or blue shark, have collagens with T_s near 55° and hydroxyproline content 10–12 per cent. Mammalian collagens (Table II) have still higher T_s values, 60° or above, and hydroxyproline content of 14 per cent or thereabouts. This suggests, as Gustavson has particularly emphasized, that the hydroxyl groups of hydroxyproline help to stabilize the native collagen structure, presumably by forming hydrogen bonds between adjacent peptide chains which serve to keep them together in an ordered configuration.

THE STRUCTURAL PATTERN—EVIDENCE FROM X-RAY AND ELECTRON MICROSCOPE STUDIES

The X-ray diffraction pattern of collagen shows a very prominent spacing along the fiber axis of 2.86 Å which is characteristic of collagen and of no other fibrous protein. There are a number of other short spacings in the wide-angle X-ray diagram, notably one at about 11 Å in dry collagen, at right angles to the fiber axis; this spacing increases up to as much as 16–17 Å in wet collagen, due to the penetration of water between the chains. If the fiber is kept stretched while its X-ray pattern is being recorded, the spacing along the fiber axis is increased from 2.86 to over 3 Å; the whole pattern becomes sharper, and a number of new reflections appear (Cowan *et al.*, 1955b). It is significant (Bear, 1955) that the general character of the X-ray pattern is characteristic of a helical structure, although the collagen helix is clearly totally different from the α -helix which we have previously discussed. We return below to the nature of the underlying structure.

In addition to this wide-angle pattern, studies of X-ray diffraction at very low angles reveal a well-defined long spacing, approximately 640 Å, along the fiber axis. This long spacing is also observed in studies of collagen with the electron microscope; preparations which are stained with such reagents as phosphotungstic acid show a series of dark bands—usually 6 or 7, but as many as 10 have been observed—within the period of 640 Å, with lightly staining regions (interbands) between the bands. The dark bands have slightly greater diameters than the interbands (Bear, 1952; Kendrew, 1954b; Schmitt *et al.*, 1955).

Many collagen fibers can be dissolved in citrate or acetate buffers at pH near 4. On readjusting the pH to 7, the dissolved molecules reassociate into fibers in an ordered fashion, so that the original collagen pattern, with its bands and interbands, may be regained. However, the reconstituted material may form banded structures with far greater periodicities than that of native collagen. These have been referred to by Schmitt *et al.* (1955), who discovered these structures, as fibrous long spacing and segment long spacing structures; the observed periodicities in the electron microscope are near 2600 Å for fibrous long spacing material and somewhat less for segment long spacing. For the details of the production of these fibers, and of their structural pattern, the reader should consult the work of Schmitt *et al.* (1955).

COLLAGEN STRUCTURE

Many hypotheses have been put forward as to the structure of collagen; most of these are now of only historical interest. A major advance in interpretation was made by Ramachandran and Kartha (1955),

who showed that many features of the wide-angle X-ray diagram could be explained by assuming a structure made up of triple chains of amino acid residues, each chain being itself a helix, the sets of chains forming a coiled coil. This proposal was modified and improved by Rich and Crick (1955, 1957), Ramachandran (1956), Cowan *et al.* (1955a), and Bear (1956). There appear to be two possible and closely related structures. These were termed Collagen I and II respectively by Rich and

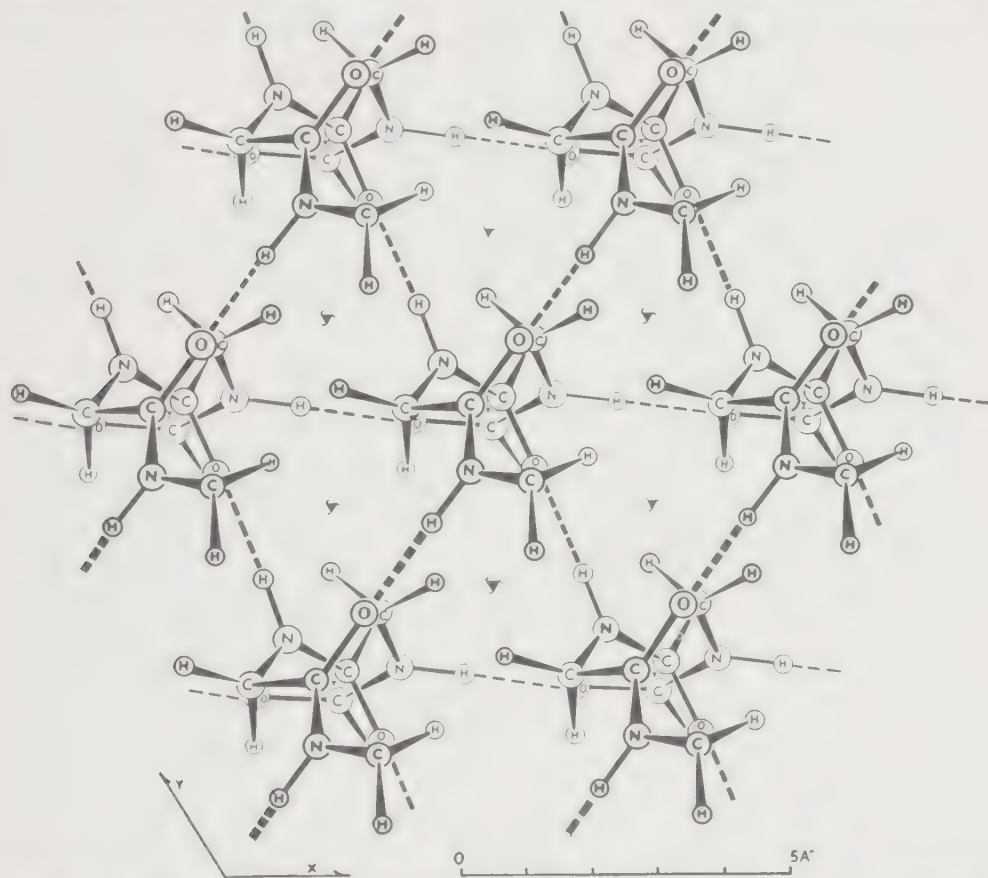


FIG. 15. The structure of polyglycine II, as proposed by F. H. C. Crick and A. Rich, *Nature* **176**, 780 (1955). A view down the screw axis is shown, showing seven polyglycine chains. Hydrogen bonds, shown as dashed lines, run in a number of directions linking neighboring chains together.

Crick, "plus" and "minus" respectively by Ramachandran, and "anti-clockwise" and "clockwise" respectively by Cowan *et al.* We shall follow most closely the discussion of Rich and Crick, which takes as a starting point a consideration of a much simpler structure—namely that of the fibrous form of polyglycine known as Polyglycine II (Bamford *et al.*, 1955). This structure, as proposed by Crick and Rich (1955), is shown from two aspects in Figs. 15 and 16. All the polyglycine chains

are parallel, each with a threefold screw axis, and are packed in a hexagonal array. Each chain is hydrogen-bonded to each of its six neighbors; the hydrogen bonds are roughly perpendicular to the screw axis; they run in several directions, in contrast to the β (pleated sheet) structures shown in Figs. 9 and 11, in which the hydrogen bonds all lie nearly in the

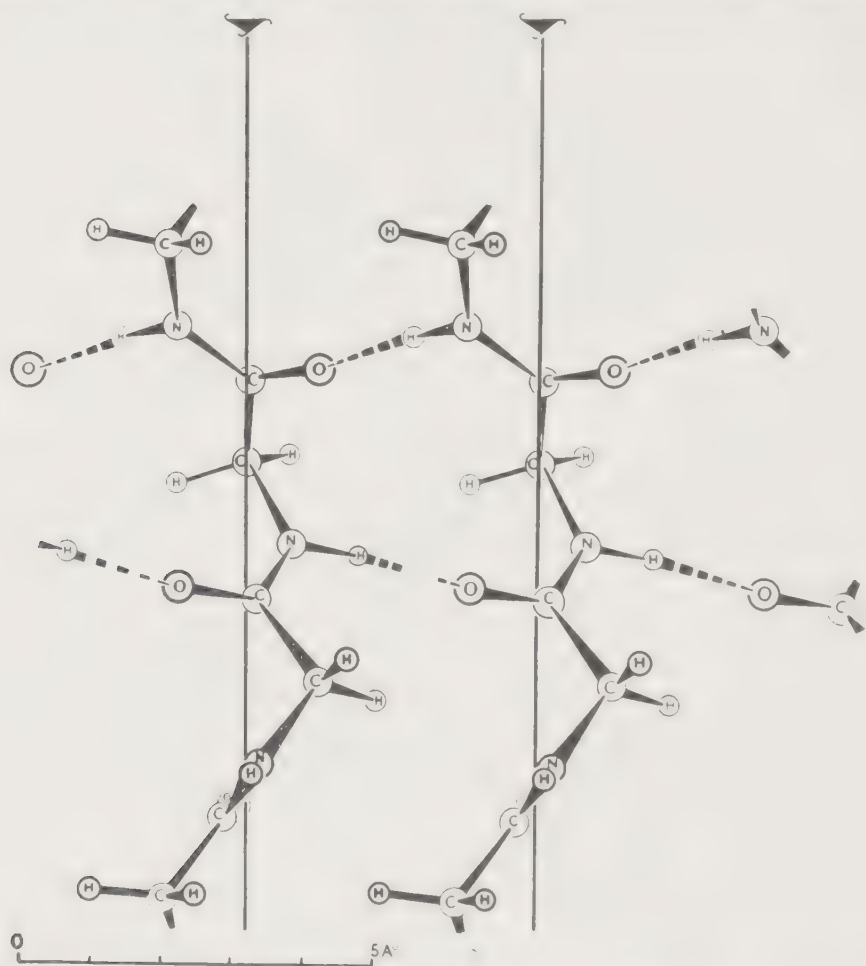


FIG. 16. Structure of polyglycine II. A projection of the structure with the screw axis vertical. The chain on the right is nearer to the reader than that on the left. Note the planar peptide groups edge-on at the bottom of the figure. The hydrogen bonds from these groups are almost perpendicular to the paper. (From F. H. C. Crick and A. Rich, 1955.)

plane of the sheet. The planar peptide groups shown for the two chains seen in Fig. 16 are inclined at about 35° to the fiber axis; the planes of the peptide groups at the lower part of the figure are very nearly normal to the plane of the paper. Moving up either one of the chains in Fig. 16, the plane of the next peptide group is rotated by 120° around the threefold screw axis, so that it lies out of the plane of the paper, as does the peptide

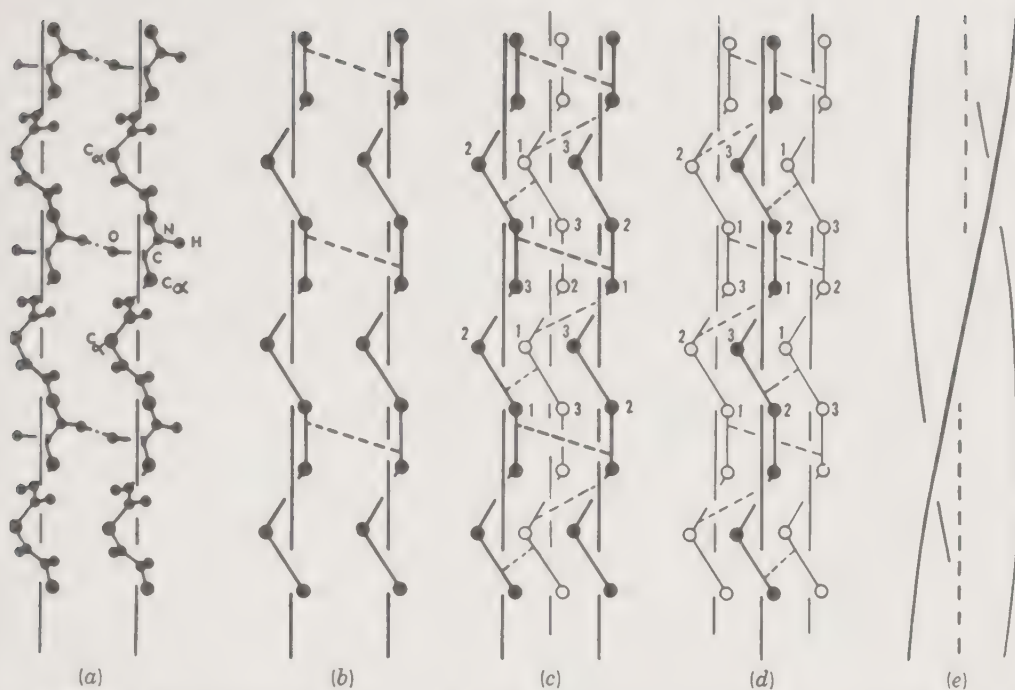


FIG. 17. Fundamental relations of peptide chains, to elucidate the general plan of the collagen structures I and II, as proposed by Rich and Crick (1957).

(a) Two polypeptide backbones shown side by side. It can be seen that each follows a helical path, having a left-handed threefold screw axis. (The axes are shown symbolically as vertical lines.) The broken lines between the two chains represent hydrogen bonds. The larger circles represent the C- α carbon atoms.

(b) A simplified version of (a), in which only the C- α carbon atoms are shown, connected by short straight lines symbolizing the peptide groups. The broken lines represent hydrogen bonding.

(c) Similar to (b), but with a third polypeptide chain added *behind* the other two. This arrangement is related to Collagen I. The numbers 1, 2, and 3 represent the three types of side-chain positions. Broken lines represent hydrogen bonds.

(d) Similar to (b), but with a third polypeptide chain added *in front* of the other two. This arrangement is related to Collagen II. The numbers 1, 2, and 3 represent the three types of side-chain positions. Broken lines represent hydrogen bonds.

(e) Showing the general way in which the structures of (c) and (d) are deformed to give the Collagen models. The solid lines represent the axes of the three polypeptide chains, which now follow gradual right-handed helices, instead of being straight and vertical. The broken line shows the common axis round which the three chains wind.

group above it. Above this the pattern is repeated. The distance in each chain per residue along the axis is approximately 3.1 Å, and the repeat distance for three residues is therefore 9.3 Å.

The basis of the proposed collagen structures I and II is shown in the five drawings of Fig. 17. On the left (Fig. 17a) is shown a pair of peptide chains, similar to those of Fig. 16, except that the coiling is left-handed instead of right-handed. The spacing in each chain, per amino acid

residue, is close to 3 Å, and the translation from one residue to the next along the chain is accompanied by a rotation of -120° , corresponding to the threefold screw axis. The screw is chosen to be left-handed so that L-prolyl residues may form part of the structure (compare the structure of poly-L-proline, Cowan and McGavin, 1955). The skeletons of these chains are shown in Fig. 17*b*. A third chain may be added either behind (Fig. 17*c*) or in front (Fig. 17*d*) of the other two chains. In either group of three chains, each is at about 5 Å from the other two. Either group of chains has a threefold screw axis running up the middle of the group of three. In any one chain, each third residue is in an identical environment; for instance, each third residue might be near the middle of the group of three chains, with the other two nearer the outside. If the three are suitably oriented, every third $-NH$ group, on the backbone of one chain, can make a hydrogen bond with every third $-CO$ group on the backbone of a neighboring chain. This holds for all three chains, since they are all equivalent. Trial by model building shows that there are only two orientations in which the three chains can be brought together to make satisfactory hydrogen bonds; that shown in Fig. 17*c* is related to Collagen I and that shown in Fig. 17*d* to Collagen II.

The actual models for collagen are obtained from these two hypothetical structures by deforming them so that their axes, instead of running straight and parallel, twist slowly around one another (Fig. 17*e*). The large helix so formed is right-handed. The threefold screw axis in the center of the group is thus deformed, its angle of rotation becoming -108° , instead of -120° . The translation per residue in any one of the peptide chains is 2.86 Å in unstretched collagen, becoming slightly greater—about 3 Å—on stretching. The rotation of -108° per residue about the axis gives a repeat after 10 residues, since the total rotation of -1080° in 10 residues is equal to 3 complete turns ($1080^\circ = 3 \times 360^\circ$); the corresponding distance along the axis is thus 28.6 Å in unstretched, or about 30 Å in stretched collagen. The significance of this tenfold repeat has been particularly pointed out by Bear (1955, 1956).

The spacing at right angles to the fiber axis, between adjacent bundles of three chains, is 15–16 Å in wet collagen, shrinking to 10.5–11 Å as the collagen is dried. The earlier observations of this spacing have already been mentioned above.

If no side chains were present, it would be equally easy to build the two structures I and II. Side chains, especially prolyl and hydroxyprolyl residues, can be added more easily to structure II than to structure I. In each chain, for either structure, the pattern repeats after each three residues; therefore there are three kinds of positions, which we denote as 1, 2, and 3 respectively. Position 1 can only be occupied by glycine, in

either structure I or II; positions 2 and 3, in structure II, may be occupied by any residue, including proline (pro) and hydroxyproline (hypro). The same is true of position 2 in structure I; position 3 for this structure, however, must be occupied by glycine, unless the structure is slightly deformed, in which case any side chain except that of valine can be fitted in.

There is evidence from amino acid residue sequence studies that the sequence gly-pro-hypro occurs frequently in collagen. This sequence would fit readily into structure II, but would also fit the deformed structure I. In the latter case the OH group of the hypro could form a hydrogen bond within the group of three chains; in structure II, the OH group would point out radially, and might form a hydrogen bond with a suitable receptor site in a neighboring group of three chains.

Evidence of a three-chain structure for collagen molecules *in solution* has been obtained by Boedtker and Doty (1956), who have studied ichthyocol of carp swim bladder, by sedimentation, viscosity, light scattering, and flow birefringence. The molecules of "native" ichthyocol have a molecular weight near 345,000, and are very long thin structures about 3000 Å long. On heat denaturation they break up into three subunits, apparently of unequal length—a finding which fits in well with the three-chain structure of collagen fibers discussed above.

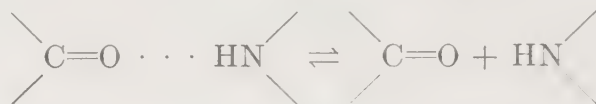
Much more remains to be done in elucidating the structure of collagen. The proposed structures I and II appear to offer a promising general interpretation of the fundamental repeating pattern; structure II appears at present more likely to be correct, although *both* structures could exist in different parts of a collagen fiber. The long range structural pattern, with the recurring bands and interbands seen in the electron microscope, is not explained by these structures, which were worked out to interpret the pattern of shorter spacings in the X-ray diagram. However, the formulation of structures I and II appears to give for the first time a really satisfactory beginning for further advances in the understanding of collagen.

The Stability of Helical Structures

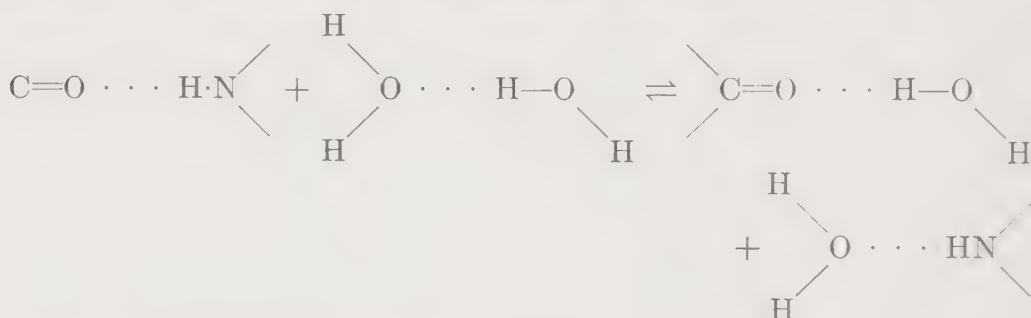
Nothing analogous to the special configuration of chains which is found in collagen has yet been recognized in other proteins. Therefore we return now to the α -helix and other related structures, which appear to have a wider range of significance for the understanding of proteins.

Since the α -helix, and the other helical structures that have been proposed for peptide chains, must be maintained in the helical configuration by $\text{—NH} \cdots \text{O=C—}$ hydrogen bonds, the stability of such structures may be greatly affected by the nature of the surrounding medium.

In a medium of low dielectric constant, incapable of hydrogen bond formation with the groups in the peptide chain, the reaction for the breakage of hydrogen bonds in the helical chain may be written



and the heat absorbed in the process is of the order of 8 kcal/mole, according to Pauling and Corey. This large energy input indicates that the bonds will not break readily. In a solvent consisting of molecules which themselves have a marked tendency to hydrogen bond formation, however, the solvent molecules may break the $\text{—NH} \cdots \text{O=C—}$ links by bonding separately with the —NH and the C=O groups. In water, for example, the type of reaction that occurs may be written



The net heat of reaction in this case, which may be considered as hydrogen bond interchange rather than hydrogen bond breakage, is naturally much less than for the other. Schellman (1955a) has estimated the heat absorbed per mole in this process as approximately 1.5 kcal/mole. This means that in aqueous solutions a helical polypeptide chain is rather likely to unfold into a more open and random type of coil. In a medium of low dielectric constant, on the other hand, it tends to remain as a helix. In media of very great hydrogen bond forming power—for instance, in concentrated urea solutions—the tendency for unfolding of the helix will be even stronger than it is in water. The C=O group and the two NH_2 groups in the urea molecule are structurally very similar to the CO and NH groups of the peptide linkage and will displace them from their intrachain hydrogen bonds in the helix, and at high concentrations of urea the bonds holding the helix structure together will tend to disintegrate. This capacity for breaking such hydrogen bonds within the structure of a peptide or protein is undoubtedly a major factor in making urea a powerful denaturing agent for proteins.

Other things being equal, a long polypeptide chain is more likely than a short one to assume the helical configuration. As may be seen

from Fig. 12, there must be three unbonded NH groups at one end of an α -helix, and three unbonded CO groups at the other end. If the chain is a short one, these unbonded terminal groups will form a large fraction of all the CO and NH groups in the peptide chain, and the forces tending to stabilize the helix, expressed per unit length of chain, will be relatively much weaker than for a long chain in which these terminal unbonded groups constitute a very small fraction of the total number of links. A thoughtful and critical survey of the whole problem of the stability of helical structures has been given by Schellman (1955b); see also Harrington and Schellman (1956).

The influence of proline residues on the stability of helical structures is important to consider. The nitrogen of a prolyl residue (Fig. 8) has no hydrogen to serve as a hydrogen bond donor; hence for each such residue in a peptide chain there must be at least one C=O group of another residue which cannot form an intrachain $\text{—C=O} \cdots \text{H—N—}$ bond. (It may, of course, form a hydrogen bond with a suitable residue of an amino acid *side chain*, if the steric relations are favorable.) Generally the effect on the continuity of the helix is more disturbing than is implied by the loss of a single hydrogen bond. In a right-handed α -helix, indeed, the insertion of a single proline residue can be accomplished—as judged by studies on models—without any far-reaching disturbance of the helical structure. One hydrogen bond would of course be eliminated, and another neighboring one would be lengthened and therefore weakened, but the continuity of the helix would not *necessarily* be sharply broken. In a left-handed α -helix, however, even a single proline residue drastically disrupts the helical sequence, and may even serve as a turning point, sharply altering the direction of subsequent portions of the peptide chain. By proper arrangement of the residues immediately adjoining the proline, it is even possible to construct models in which the helix reverses direction and loops back on itself through 180° . The making of such a loop of course involves the breaking of several hydrogen bonds in the residues immediately adjoining the turning point. (See the discussion by Low and Edsall, 1956, and the drawings on pp. 412 and 413 of their article.)

The Significance of Disulfide Bonds; Cross-Linkages and Loops in Peptide Chains

Disulfide bonds, in a protein structure, may serve to link two different peptide chains together, or to join two different segments of the same peptide chain, thus producing a loop. In insulin (Fig. 5) both types of disulfide linkage are found. In ribonuclease, there is only a single peptide chain, 124 residues long, and the four disulfide bridges present require that the chain must turn a corner, and reverse direction, in several places.

It has sometimes been suggested that prolyl residues form the turning points when such reversal of direction occurs, as we have indicated in the paragraph above. There are four prolyl residues in ribonuclease, but two of these are located in the C-terminal "tail" of the molecule, beyond the disulfide bond system, so they cannot function in the formation of internal loops. The other two would not be sufficient to provide all the required loops, so that some other mechanism for the reversal of chain direction must also be operative, if indeed the loops involve proline at all. Some possibilities have been suggested by Lindley (1955), who has shown for instance by the use of space models that the sequence Ser-Gly-X-X-GluNH₂-Asp-Gly could serve as a region permitting a 180° turn in a left-handed α -helix. (The two X's in this formula may be any kind of amino acid residue.) The model with the 180° turn can be made from a straight-chain α -helix by the rupture of four —NH · · · OC bonds, but in the final proposed structure five new —NH · · · OC bonds are formed between the glutamine and aspartyl side-chain amide or carboxylate groups and the main-chain C=O and NH groups; also one OH · · · OC bond from the seryl residue. These hydrogen bonds should tend to stabilize the structure. The two small glycyl residues, at the appropriate places in the sequence, are required to permit sufficiently close fitting of the two portions of the chain around the bend. It should be emphasized that the sequence described above has not yet been demonstrated in any protein. It is merely one of a number of hypothetical possibilities, which may have suggestive value for future research.

Possible Stereochemistry of Insulin

Cross-linking disulfide linkages may serve either to stabilize or to inhibit the formation of an α -helix. If two chains lie side by side, owing to such cross-linkages, as is probably true for the A and B chains of insulin, then the tendency of the chains to assume a helical configuration is probably increased. On the other hand; the intrachain disulfide bond between residues 6 and 11 in the A chain involves a twisting of the chain, which is incompatible with the formation of an α -helix for the residues between 6 and 11. A critical examination of the stereochemistry of insulin has been made by Lindley and Rollett (1955). They have found that it is possible to build a space model of the insulin molecule which is in accordance with the sequences and structure indicated in Table IV and Fig. 5, and which consists for the most part of an α -helix. The B chain in their model is a right-handed α -helix. The A chain is a left-handed helix between residues 1 and 9, and a right-handed helix from residues 9 to 21. The region in the A chain between residues 6 and 11 forms a sort of figure of eight which is required by this change of hand in

the coil. Low (see Low and Edsall, 1956) has independently constructed another space model for insulin, using a similar change of hand from a left-handed segment for residues A₁ to A₈, changing to a right-handed segment from A₁₁ to A₂₁. The B chain of her model is a left-handed helix. It still remains to be determined whether either of these models is at all close to an actual representation of the stereochemistry of insulin. Certainly there is no proof as yet that the α -helix plays any part in the actual structure of the molecule. One suggestive piece of evidence, however, comes from the X-ray studies of Low (1952) on crystals of acid insulin sulfate, which are orthorhombic. The X-ray data give suggestive evidence of rodlike chain structures running parallel to the *a*-axis of the unit cell, which is 44 to 44.5 Å in length. It is natural to infer that the chains which appear to be present are actually peptide chains, and it appears more than a coincidence that the B chain—the longer of the two peptide chains in Fig. 5—is composed of 30 residues. Since the length of an α -helix is 1.47 to 1.50 Å per residue, the length of the B chain would be 44.1 to 45.0 Å on this basis, and it would fit into the unit cell very neatly. Further details of the possible packing arrangements have been discussed by Low (1952, 1953).

Another elegant method of study is that of determining the rate of deuterium exchange between a protein molecule and the surrounding aqueous medium (Linderstrøm-Lang, 1955; Hvidt and Linderstrøm-Lang, 1955). In simple organic compounds all hydrogen in amino, carboxyl, hydroxyl, amide, and guanidino groups is instantaneously exchangeable for deuterium, or vice versa. The same is true of the nitrogen-linked hydrogen in peptide linkages, and in imidazole groups. In a protein, however, if some of these groups are held by strong internal hydrogen bonding, they may exchange very slowly or not at all. In the A and B chains of beef insulin, there are 91 potentially exchangeable hydrogens—48 hydrogens of peptide linkages (the one prolyl residue in the B chain contains no peptide hydrogen), and 43 in the A and B side chains, if we consider insulin in acid solution in which each carboxyl group carries one attached proton and each free amino group, being positively charged, carries three. Native insulin at pH 3 and 0° exchanges 60 of these 91 atoms with a half-time below one minute; three other sets of groups with roughly 6, 15, and 8 hydrogen atoms, respectively, exchange with half-times of 1.4, 20, and ∞ hours, respectively. It seems probable that the more stable hydrogen atoms are those in the peptide links located in the central portions of the A and B chains, between the two interchain disulfide bonds. The loose ends of these chains would probably exchange their hydrogens more readily, as would the peptide hydrogens in the loop of the A chain produced by the intrachain disulfide link between A₆ and

A₁₁. These findings are not proof that the more slowly exchangeable hydrogens are held in an α -helix structure, but the existence of some kind of hydrogen-bonded structure in part of the insulin molecule is certainly indicated. We may note that this bonding is much weakened by raising the temperature to 38° or by dissolving the insulin in 5.2 *M* urea. Either of these treatments greatly increases the exchange rates for the less readily exchangeable hydrogens. Neither treatment, however, destroys the biological activity of insulin, or its ability to crystallize under suitable conditions. Thus the "loosening" of structure produced by warming or by urea treatment must be reversible.

Stereochemistry and Enzyme Activity of Ribonuclease

The structural formula of ribonuclease, shown in Fig. 6, indicates immediately that the single peptide chain must make several loops in order to accommodate the constraints imposed by the four disulfide linkages. There is other evidence that the molecule is highly compact and not far from spherical in its general shape. The greatest dimension of the unit cell of the crystal is only 52 Å, and this means that the largest dimension of the molecule is in all probability smaller than this. Moreover, the motion of the ribonuclease molecule in solution, either in free diffusion or during sedimentation in an ultracentrifuge, is very much what would be expected from a spherical molecule of the same molecular weight, with a small amount of water attached; an extended polypeptide chain, either in the form of a rigid helix or of a random coil of irregular shape, would experience far greater resistance to its motion through the liquid than the actual molecule does. (The theory of such sedimentation and diffusion studies is discussed in Volume II.)

How much of the peptide chain of ribonuclease may be in the form of an α -helix in the native molecule is very hard to tell. From the structural formula (Fig. 6) one may calculate that there are 243 exchangeable hydrogens at pH 4.7. Of these, 123 are peptide chain hydrogens. Hvidt (1955)—see also Anfinsen and Redfield (1956, p. 70)—has found that all 243 atoms exchange instantaneously with deuterium if the disulfide links have been previously broken by oxidation with performic acid or if the native ribonuclease is dissolved in 2.5 *M* guanidinium chloride. In the native molecule in water at 0°, however, only about 185 of the hydrogens exchange so promptly. Here again, as in insulin, rise of temperature increases the rate of exchange for the less readily exchangeable hydrogens. In the words of Hvidt, this suggests "some kind of hydrogen-bonded folded structure in the molecule, shielding some part of the hydrogen atoms from instantaneous exchange." This bonding can readily be broken, however, as in 8 *M* urea or 2.5 *M* guanidinium chloride, without destroy-

ing the activity of ribonuclease as an enzyme. We may remark also that enzymic activity is fully retained if the last three residues—Ala, Ser, and Val—at the C-terminal end of the molecule are removed by carboxypeptidase. Removal of the fourth residue (Asp), however, results in complete loss of activity. The problem of the nature of the active center of the molecule is thoughtfully discussed by Anfinsen and Redfield (1956).

Ribonuclease contains six tyrosine hydroxyl groups. Three of these are apparently free to ionize, but the other three are rather tightly hydrogen-bonded to some other element of the structure. The evidence for this feature of the structure is discussed in Chapter 9.

Helical and Nonhelical Regions in Globular Proteins

There is now some evidence that only a very small fraction of the peptide chain of ribonuclease, even in the undenatured molecule in water, is coiled in a regular helix. This evidence comes from studies of the dispersion of optical rotation—that is, the variation of the rotation with the wavelength of the light used—in synthetic polypeptides and proteins. Synthetic polypeptides, such as poly- γ -benzyl-L-glutamate, which can be dissolved in a wide variety of solvents, exist in the helical form in chloroform or dioxane, which have little tendency to form hydrogen bonds with C=O or N—H groups, but unfold into a randomly coiled form in the presence of solvents like dichloroacetic acid, with a great power of hydrogen bond formation (Doty *et al.*, 1956). The transition from the random coil to the helix is accompanied by a rise in the optical rotation in the visible region of the spectrum, from markedly negative to distinctly positive values. Also the form of the curves for optical rotation as a function of wavelength is profoundly changed (Yang and Doty, 1957) (Fig. 18). This effect of helix formation on optical rotation is to be expected on theoretical grounds (Moffitt and Yang, 1956; Moffitt, 1956; Fitts and Kirkwood, 1956). Moreover, the sign of the observed change associated with helix formation indicates that the helices formed from peptides containing L-amino acid residues are probably right-handed, not left-handed (see Fig. 12). Studies on poly-L-alanine, both by X-ray diffraction (Elliott and Malcolm, 1956) and by optical rotation (Elliott *et al.*, 1956), also indicate that, when in the α -form, it exists as a right-handed α -helix. These conclusions should not yet be taken as final, but it does seem clear that one helical form is more stable than the other for such L-polypeptides, and that this is probably the right-handed form.

Yang and Doty (1957) have examined a number of proteins by the optical rotatory dispersion technique and have attempted to infer, from the curves obtained, the proportion of the peptide chain (or chains) which is in the form of a helix. These values range from 15% helix in ribonuclease

to about 45% in ovalbumin and serum albumin. The preferred sense of twist in the helices of the proteins is the same as in the synthetic polypeptides of L-amino acids, i.e., presumably right-handed. The case of insulin is more complicated. If some of the A chain is a right-handed helix, passing over to a left-handed helix within the intrachain disulfide

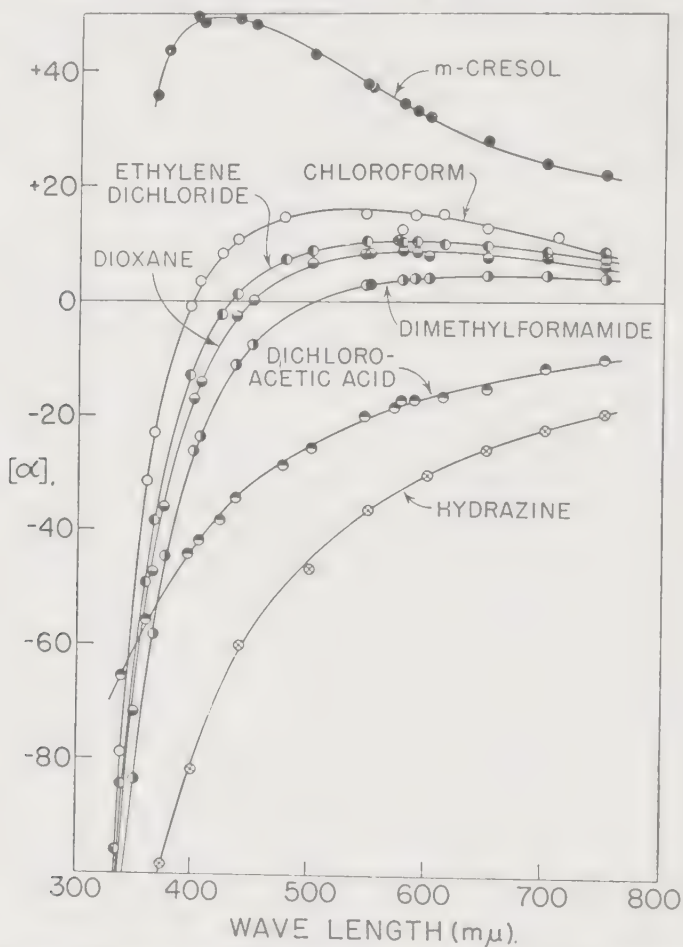


FIG. 18. Rotatory dispersion of poly- γ -benzyl-L-glutamate (MW = 130,000) in several solvents. The structure of the polymer is that of a random coil in hydrazine and dichloroacetic acid, and an α -helix in the other solvents shown. (From Yang and Doty (1957), p. 763.)

loop, as suggested above (p. 126) the right- and left-handed helical contributions to the optical rotation will partially cancel one another out. If this were the actual structure of the A chain, the amount of helix in the insulin structure would be higher than that estimated from the optical rotation. Allowing for this, the data of Yang and Doty for insulin may be compatible with a high helical content for this protein in solution.

Cohen and Szent-Györgyi (1957) have made similar studies on the muscle proteins tropomyosin, myosin, and the light and heavy meromyo-

sins derived from myosin by gentle enzymatic treatment. They also studied the blood protein fibrinogen, the precursor of the fibrin clot. Tropomyosin is more than 90% helical, according to the criteria of optical rotation used by Yang and Doty. The other elongated proteins studied by Cohen and Szent-Györgyi all show a higher helical content than do most globular proteins, although not as high as tropomyosin. The sense of twist in the helix is the same in these proteins as in the other peptides and proteins studied by Yang and Doty. In a later note, Szent-Györgyi and Cohen (1957) point out an inverse relation between proline content of proteins and tendency to helix formation; proteins inferred from optical rotation studies to be more than 50% helical almost always contain less than 3% proline. Collagen, with its very high proline content, is a special case, since the helix of the collagen fiber is entirely different from an α -helix.

Further tests of these conclusions by other techniques are certainly desirable, but these results emphasize the inference that, in globular protein molecules, we should not expect to find so simple a pattern as a set of packed α -helices, merely interrupted by occasional loops in the chain where the helix structure is necessarily broken. It is probable that the "irregular"—that is, nonhelical—portions of the peptide chains in globular proteins are often more extensive than the "regular" regions. This irregularity does not mean randomness, however; the beautiful X-ray diffraction patterns given by crystalline proteins clearly imply that the molecules are highly orderly structures. Rather we may suspect that the folds and twists of the peptide chains in globular proteins are specific and capable of great individual variation from molecule to molecule, rather than following any one simple kind of repeating pattern, such as the α -helix.

X-ray diffraction studies on certain globular proteins have now given a vast amount of information on the positions and intensities of the X-ray reflections. As yet, very little of this information has been translated into data defining the structure of the molecules. Recently progress has greatly accelerated in this field, owing largely to the technique of incorporating heavy atoms in such crystals, in a few specifically reactive places in the protein molecules, and observing the changes of intensity in the X-ray reflections that result from these chemical modifications. From these changes it is possible to determine the coordinates of the added heavy atoms, within the unit cell of the crystal, and thence to infer, by a series of elaborate but quite straightforward steps, the distribution of electron density—that is, of the atoms—in the whole molecule. At the moment, no three-dimensional picture of such a distribution has been achieved, but only projections of the distribution along certain axes of the unit

cell for hemoglobin and myoglobin (Crick and Kendrew, 1957). The detailed pattern of the molecule has still not been visualized from such measurements. For instance, Kendrew and Parrish (1957), after a prolonged study of sperm whale myoglobin, inferred the general shape of the molecule to be something rather like a triaxial ellipsoid, with axes approximately $25 \times 34 \times 42$ Å. It is known from chemical studies that there is only one peptide chain per molecule, approximately 142 residues long, and one flat heme group with the iron atom at its center. There are indications that the prevailing orientation of the peptide chain is parallel to the longest axis of the molecule, but the chain must loop back and forth on itself several times in order to fit into so compact a molecule. The angle between this inferred direction of the peptide chains and the plane of the heme group is 40 to 50°, from optical measurements on the orientation of preferred absorption of polarized light by the heme group. This rather meager description roughly summarizes structural knowledge concerning myoglobin at present, but every prospect exists that the analysis of X-ray measurements will soon lead to a far deeper and more detailed picture of the structure. Similarly Bragg *et al.* (1954) and Bragg and Perutz (1954) have evolved a picture of hemoglobin—four times as big as myoglobin, and with four heme groups—as a spheroid $53 \times 53 \times 71$ Å in the case of the hydrated molecule, and $45 \times 45 \times 65$ Å for the dry molecule. This simple description of the shape indicates only the rough outlines—a detailed picture would undoubtedly show a very varied pattern, with many knobs and hollows. The achievement of that detailed picture, for certain proteins at least, seems now almost within our grasp. As it is achieved, rational explanations of such phenomena as enzyme and antibody specificity should rapidly be attained, and a far deeper understanding of such phenomena as the action of drugs on living tissues should follow. Further progress in the field of protein chemistry should indeed be crucial for the whole of biology and medicine.

GENERAL REFERENCES

- Anson, M. L., Bailey, K., and Edsall, J. T. (eds.). *Advances in Protein Chemistry*, Academic Press, New York. Eleven volumes have appeared through 1956. A number of references to articles in this series are given below.
- Cohn, E. J., and Edsall, J. T. (1943). "Proteins, Amino Acids and Peptides," Reinhold Publishing Corp., New York.
- Neurath, H., and Bailey, K. (eds.) (1953-54). "The Proteins," Vols. IA, IB, IIA, and IIB, Academic Press, New York. Several references to individual chapters in this book are given below.
- Springall, H. D. (1954). "The Structural Chemistry of Proteins," Butterworth's Scientific Publications, London, and Academic Press, New York.
- Vickery, H. B., and Schmidt, C. L. A. (1931). The history of the discovery of the amino acids, *Chem. Revs.* **9**, 169.

"Fibrous Proteins and Their Biological Significance" (1955). *Symposia of the Society for Experimental Biology*, Number IX, Academic Press, New York.

SPECIAL REFERENCES

- Anfinsen, C. B., and Redfield, R. R. (1956). *Advances in Protein Chem.* **11**, 1.
- Astbury, W. T., and Street, A. (1931). *Trans. Roy. Soc.* **A230**, 75.
- Astbury, W. T., and Woods, H. J. (1933). *Trans. Roy. Soc.* **A232**, 333.
- Astbury, W. T., and others (1953). A discussion on the structure of proteins, *Proc. Roy. Soc.* **B141**, 1-102 incl.
- Bamford, C. H., Brown, L., Cant, E. M., Elliott, A., Hanby, W. E., and Malcolm, B. R. (1955). *Nature* **176**, 396.
- Bear, R. S. (1952). The structure of collagen fibrils, *Advances in Protein Chem.* **7**, 69.
- Bear, R. S. (1955). Fibrous proteins and their biological significance, *Symposia Soc. Exptl. Biol.* **9**, 97.
- Bear, R. S. (1956). *J. Biophys. Biochem. Cytol.* **2**, 363.
- Bell, P. H. (1954). *J. Am. Chem. Soc.* **76**, 5566.
- Benesch, R. E., Lardy, H. A., and Benesch, R. (1955). *J. Biol. Chem.* **216**, 663.
- Block, R. J., Durrum, E. L., and Zweig, G. (1955). "A Manual of Paper Chromatography and Paper Electrophoresis," Academic Press, New York.
- Boedtker, H., and Doty, P. (1956). *J. Am. Chem. Soc.* **78**, 4267.
- Bragg, W. L. (1931). "The Crystalline State," Vol. 1, General Survey, H. G. Bell and Sons, Ltd., London.
- Bragg, Sir L., and Perutz, M. F. (1954). *Proc. Roy. Soc.* **A225**, 315.
- Bragg, Sir L., Howells, E. R., and Perutz, M. F. (1954). *Proc. Roy. Soc.* **A222**, 33.
- Bricas, E., and Fromageot, C. (1953). *Advances in Protein Chem.* **8**, 1.
- Bromer, W. W. *et al.* (1957). *J. Am. Chem. Soc.* **79**, 2794-2810. A series of five papers.
- Brown, H., Sanger, F., and Kitai, R. (1955). *Biochem. J.* **60**, 556.
- Bunn, C. W. (1945). "Chemical Crystallography," Oxford: Clarendon Press.
- Cohen, C., and Szent-Györgyi, A. G. (1957). *J. Am. Chem. Soc.* **79**, 248.
- Cowan, P. M., and McGavin, S. (1955). *Nature* **176**, 501.
- Cowan, P. M., McGavin, S., and North, A. C. T. (1955a). *Nature* **176**, 1062.
- Cowan, P. M., North, A. C. T., and Randall, J. T. (1955b). Fibrous proteins and their biological significance, *Symposia Soc. Exptl. Biol.* **9**, 115.
- Crammer, J. L., and Neuberger, A. (1943). *Biochem. J.* **37**, 302.
- Crick, F. H., and Kendrew, J. C. (1957). *Advances in Protein Chem.* **12** (in press).
- Crick, F. H. C., and Rich, A. (1955). *Nature* **176**, 780.
- Donohue, J. (1953). *Proc. Natl. Acad. Sci. (U.S.)* **39**, 470.
- Doty, P., Bradbury, J. H., and Holtzer, A. M. (1956). *J. Am. Chem. Soc.* **78**, 947.
- du Vigneaud, V. (1956). *Science* **123**, 967.
- du Vigneaud, V., Ressler, C., Swan, J. M., Roberts, C. W., and Katsoyannis, P. G. (1954). *J. Am. Chem. Soc.* **76**, 3115.
- Edsall, J. T. (1947). *Advances in Protein Chem.* **3**, 383.
- Elliott, A., and Malcolm, B. R. (1956). *Nature* **178**, 912.
- Elliott, A., Hanby, W. E., and Malcolm, B. R. (1956). *Nature* **178**, 1170.
- Fitts, D. D., and Kirkwood, J. G. (1956). *Proc. Natl. Acad. Sci. (U.S.)* **42**, 33; *J. Am. Chem. Soc.* **78**, 2650.
- Geschwind, I. I., Li, C. H., and Barnafi, L. (1957a). *J. Am. Chem. Soc.* **79**, 620.
- Geschwind, I. I., Li, C. H., and Barnafi, L. (1957b). *J. Am. Chem. Soc.* **79**, 1003.
- Green, A. A., and Hughes, W. L. (1955). In "Methods in Enzymology" (S. Colowick and N. O. Kaplan, eds.), Vol. I, p. 67, Academic Press, New York.

- Green, N. M., and Neurath, H. (1954). Proteolytic enzymes, in "The Proteins" (H. Neurath and K. Bailey, eds.), Vol. IIB. Chapter 24.
- Greenstein, J. P. (1954). *Advances in Protein Chem.* **9**, 121.
- Gustavson, K. H. (1956). "The Chemistry and Reactivity of Collagen." Academic Press, New York.
- Harfenist, E., and Craig, L. C. (1952). *J. Am. Chem. Soc.* **74**, 3083, 3087.
- Harrington, W. F., and Schellman, J. A. (1956). *Compt. rend. trav. lab. Carlsberg Sér. chim.* **30**, 21.
- Harris, J. I., and Roos, P. (1956). *Nature* **178**, 90.
- Harris, J. I., Sanger, F., and Naughton, M. A. (1956). *Arch. Biochem. and Biophys.* **65**, 427.
- Hirs, C. H. W., Stein, W. H., and Moore, S. (1956). *J. Biol. Chem.* **221**, 151; **219**, 611, 623; **221**, 143.
- Howard, K. S., Shepherd, R. G., Eigner, E. A., Davis, D. S., and Bell, P. H. (1955). *J. Am. Chem. Soc.* **77**, 3419.
- Hughes, W. L. (1946). *J. Am. Chem. Soc.* **69**, 1836.
- Hvidt, A. (1955). *Biochim. et Biophys. Acta* **18**, 306.
- Hvidt, A., and Linderstrøm-Lang, K. (1955). *Compt. rend. trav. lab. Carlsberg Sér. chim.* **29**, 385.
- Ingram, V. M. (1955). *Biochem. J.* **59**, 653.
- Kauzmann, W. (1954). In "The Mechanism of Enzyme Action" (W. D. McElroy and B. Glass, eds.), p. 70. Johns Hopkins Press, Baltimore, Maryland.
- Kendrew, J. C. (1954a). *Progr. Biophys. and Biophys. Chem.* **4**, 244.
- Kendrew, J. C. (1954b). In "The Proteins" (H. Neurath and K. Bailey, eds.), Vol. IIB, Chapter 23, Academic Press, New York.
- Kendrew, J. C., and Parrish, R. G. (1957). *Proc. Roy. Soc.* **A238**, 305.
- Kendrew, J. C., and Perutz, M. F. (1949). In "Haemoglobin" (F. J. W. Roughton and J. C. Kendrew, eds.), p. 161. Butterworth's Scientific Publications, London, and Interscience Publishers, New York.
- Lederer, E., and Lederer, M. (1953). "Chromatography." Elsevier, Amsterdam. (Second ed. 1957)
- Levy, A. L. (1954). *Nature* **174**, 126.
- Li, C. H. (1956). *Advances in Protein Chem.* **11**, 101; (1957). *ibid.* **12** (in press).
- Li, C. H., Geschwind, I. I., Cole, R. D., Raache, I. D., Harris, J. I., and Dixon, J. S. (1955). *Nature* **176**, 686.
- Linderstrøm-Lang, K. (1955). In "Peptide Chemistry: Report of a Symposium Held by the Chemical Society, March 30, 1955," The Chemical Society, London.
- Lindley, H. (1955). *Biochim. et Biophys. Acta* **18**, 194.
- Lindley, H., and Rollett, J. S. (1955). *Biochim. et Biophys. Acta* **18**, 183.
- Lonsdale, K. (1949). "Crystals and X-Rays," Van Nostrand Co., Princeton, New Jersey.
- Low, B. W. (1952). *Nature* **169**, 955.
- Low, B. W. (1953). In "The Proteins" (H. Neurath and K. Bailey, eds.), Vol. IA, Chapter 4, Academic Press, New York.
- Low, B. W., and Edsall, J. T. (1956). In "Currents in Biochemical Research 1956" (D. E. Green, ed.), p. 378, Interscience Publishers, New York.
- Low, B. W., and Grenville-Wells, H. J. (1953). *Proc. Natl. Acad. Sci. (U.S.)* **39**, 785.
- Marsh, R. E., Corey, R. B., and Pauling, L. (1955). *Biochim. et Biophys. Acta* **16**, 1.
- Martin, A. J. P. (1950). *Ann. Rev. Biochem.* **19**, 517.
- Moffitt, W. (1956). *Proc. Natl. Acad. Sci. (U.S.)* **42**, 736.
- Moffitt, W., and Yang, J. T. (1956). *Proc. Natl. Acad. Sci. (U.S.)* **42**, 596.
- Moore, S., and Stein, W. H. (1948). *J. Biol. Chem.* **176**, 367.

- Moore, S., and Stein, W. H. (1951). *J. Biol. Chem.* **192**, 663.
- Moore, S., and Stein, W. H. (1952). *Ann. Rev. Biochem.* **21**, 521.
- Moore, S., and Stein, W. H. (1954a). *J. Biol. Chem.* **211**, 893.
- Moore, S., and Stein, W. H. (1954b). *J. Biol. Chem.* **211**, 907.
- Moore, S., and Stein, W. H. (1956). *Advances in Protein Chem.* **11**, 191.
- Neuberger, A. (1948). *Advances in Protein Chem.* **4**, 297.
- Pauling, L. (1953). In "Les Proteines: Rapports et Discussions" (R. Stoops, ed.), p. 63, Institut International de Chimie Solvay, Brussels.
- Pauling, L., and Corey, R. B. (1951). *Proc. Natl. Acad. Sci. (U.S.)* **37**, 205-285, 729.
- Pauling, L., and Corey, R. B. (1953). *Proc. Natl. Acad. Sci. (U.S.)* **39**, 253.
- Perlmann, G. (1955). *Advances in Protein Chem.* **10**, 1.
- Ramachandran, G. N. (1956). *Nature* **177**, 710.
- Ramachandran, G. N., and Kartha, G. (1955). *Nature* **176**, 593.
- Redfield, R. R., and Anfinsen, C. B. (1956). *J. Biol. Chem.* **221**, 385.
- Rhinesmith, H. S., Schroeder, W. A., and Pauling, L. (1957). *J. Am. Chem. Soc.* **79**, 4682.
- Rich, A., and Crick, F. H. C. (1955). *Nature* **176**, 915.
- Rich, A., and Crick, F. H. C. (1957). In Symposium on Gelatin and Glue, Cambridge, England, July, 1957.
- Riggs, A. F., and Wolbach, R. A. (1956). *J. Gen. Physiol.* **39**, 585.
- Robertson, J. M. (1953). "Organic Crystals and Molecules," Cornell University Press, Ithaca, New York.
- Ryle, A. P., and Anfinsen, C. B. (1957). *Biochim. et Biophys. Acta* **24**, 633.
- Ryle, A. P., and Sanger, F. (1955). *Biochem. J.* **60**, 535.
- Ryle, A. P., Sanger, F., Smith, L. F., and Kitai, R. (1955). *Biochem. J.* **60**, 541.
- Sanger, F. (1952). *Advances in Protein Chem.* **7**, 1.
- Sanger, F. (1956). In "Currents in Biochemical Research 1956" (D. E. Green, ed.), p. 434, Interscience Publishers, New York.
- Sanger, F., et al. (1951-55). A series of papers on the structure of insulin: *Biochem. J.* **49**, 463, 481 (1951); **53**, 353, 366 (1953); **59**, 509 (1954); **60**, 535, 541, 556 (1955).
- Schellman, J. A. (1955a). *Compt. rend. trav. lab. Carlsberg Sér. chim.* **29**, 230-259.
- Schellman, J. A. (1955b). *Compt. rend. trav. lab. Carlsberg. Sér. chim.* **29**, 223-229.
- Schmitt, F. O., Gross, J., and Highberger, J. H. (1955). Fibrous proteins and their biological significance, *Symposia Soc. Exptl. Biol.* **9**, 148.
- Snell, E. E. (1946). *Advances in Protein Chem.* **2**, 85.
- Stein, W. H., Kunkel, H. G., Cole, R. D., Spackman, D. H., and Moore, S. (1957). *Biochim. et Biophys. Acta* **24**, 640.
- Steinberg, D., and Mihalyi, E. (1957). *Ann. Rev. Biochem.* **26**, 373.
- Straessle, R. (1954). *J. Am. Chem. Soc.* **76**, 3138.
- Szent-Györgyi, A. G., and Cohen, C. (1957). *Science* **126**, in press.
- Taylor, J. F. (1953). In "The Proteins" (H. Neurath and K. Bailey, eds.), Vol. IA, Chapter 1, Academic Press, New York.
- Thompson, E. O. P., and Thompson, A. R. (1955). *Fortschr. Chem. org. Naturstoffe* **12**, 270.
- Tristram, G. R. (1953). In "The Proteins" (H. Neurath and K. Bailey, eds.), Vol. IA, Chapter 3, Academic Press, New York.
- Turba, F. (1954). "Chromatographischen Methoden in der Protein-Chemie," Springer, Berlin.
- Wald, G. (1953). *Federation Proc.* **12**, 606.
- Wald, G. (1954). *Am. Scientist* **42**, 73.
- Yang, J. T., and Doty, P. (1957). *J. Am. Chem. Soc.* **79**, 761.

Chapter 4

Thermodynamics

Introduction

Thermodynamics, as its name implies, is a subject which has to do with heat and temperature. In this way it is to be distinguished from classical mechanics. Owing to the fact that temperature is a universal attribute of all actual substances, thermodynamics is of enormously wide applicability. Not all thermal phenomena, however, fall within its scope. For example, the conduction of heat, the theory of which has been so highly developed since it took its beginnings over a century ago at the hands of Fourier, has little or nothing to do with thermodynamics. This represents the fact that thermodynamics is primarily limited to the study of equilibria, though it also has much to say, at a very general level, concerning the direction of processes in systems not at equilibrium, and provides criteria for irreversible processes.

The early development of thermodynamics in the first half of the nineteenth century was closely associated with the study of heat engines. It soon became apparent, however, that the subject had far wider applications, and its principles rapidly assumed an important role in all branches of physical science. It became, among other things, a principal source of the ideas of physical chemistry. Indeed, the classical physical chemistry dealing with equilibria is to be regarded as a branch of thermodynamics, and the term chemical thermodynamics has been introduced to refer to it. This subject comes to us by direct succession from the work of Willard Gibbs, who saw it as a whole not only for the first time but also from a height which has perhaps never been regained. It is the purpose of this chapter to give as far as possible a brief account of the main principles of chemical thermodynamics, which will serve as a review for a student who has already some acquaintance with it and as an introduction to it for one who has none.

Systems and Phases

Because of the abstractness and generality of its approach, in which the effort is to arrive at comprehensive principles independent of the particular characteristics of any given situation, thermodynamics has

need of a term to refer to the most general type of physical situation. The word *system* is used for this purpose. It is applied to any part of the physical universe under consideration, regardless of how the part is distinguished from the rest. Thus, for example, we may speak of a body of gas in a closed vessel as a system, or we may use the term to include both the gas and the vessel. We may speak of the human body as a system, or any part of it under consideration, such as the blood, or a muscle. The ocean might be referred to in the same way, or the term might be applied to the earth itself, or, in the limiting case, the entire physical universe.

A system may or may not involve a fixed amount of material. Thus, in the example of the gas, the system is composed of a permanent population of molecules. In the case of the blood there are constant exchanges of matter between the system and its surroundings, e.g., O_2 , CO_2 , H_2O . A system comprising a fixed amount of matter—that is, one which carries out no exchange of matter with its surroundings—is called a *closed system*. A closed system which does not interact with its surroundings in any way, as by the absorption or liberation of heat or the performance of mechanical work, is said to be *isolated*. It should be realized, however, that the concept of an isolated system represents an abstraction; for example, no thermal insulation is ever perfect. But the greater the extent of a system, the more susceptible it is of isolation. This follows from the two facts that exchanges of matter and energy can occur only through the bounding surface and that the amounts of matter and energy contained within the system depend on its volume, the ratio of surface to volume diminishing with absolute size.

A system may or may not be of uniform composition, that is to say homogeneous, down to molecular dimensions. Homogeneity prevails in the case of the gas but not in the case of the blood, where there are present both cells and plasma. In the former case, the system is said to consist of a single phase. Otherwise, it is a multiphase system. It should be emphasized, however, that the number of phases in a system is equal to the number of different *homogeneous* regions which it contains but does not depend on the physical continuity of these regions. Thus a system composed of water and cracked ice is a two-phase system, although the ice is subdivided into a large number of separate fragments, all alike as regards molecular composition.

Particularly in the study of some biochemical systems, it may become difficult to decide whether a given system is or is not homogeneous, in the sense of being a single phase. Is a solution of a single pure protein, such as egg albumin in water, to be regarded as a single phase, even though each protein molecule is several thousand times as large as a water molecule? If this is to be considered a one-phase system, is the same true

of a solution of tobacco mosaic virus particles, each of which is about a thousand times as large as an albumin molecule? If we consider particles of progressively larger and larger size, we must eventually consider them as forming a separate phase. No simple arbitrary answer to these questions is possible. It is found in practice, however, that solutions containing very large molecules can often be well described in terms of the same thermodynamic relations that suffice for solutions of small molecules.

The description of a one-phase system involves the specification of various properties, such as pressure, volume, temperature, surface, and the masses of the various components. Certain of these—for example, pressure and temperature—are independent of the total mass of the system and are called intensive properties. Others, like volume and surface, depend on its mass and are for that reason called extensive properties. If we are concerned only with the internal state of a system, which is, of course, independent of its total mass, we may replace the extensive properties by corresponding intensive ones, obtained by referring them to unit mass; e.g., for volume we may substitute specific volume, which is the reciprocal of density, or for the mass of a component we may substitute its weight fraction. In the description of a multiphase system, it is necessary not only to give the values of the variables which define the state of each individual phase, but to specify also the nature of the contact between the phases, as, for example, whether they are separated by semipermeable membranes, and if so, of what character.

Temperature

The concepts of volume, surface, and pressure are geometrical and mechanical in nature and have exact definitions in terms of the fundamental mechanical units of mass, length, and time. These determine their “dimensions.” The concept of temperature, on the other hand, lies outside the scope of mechanics and belongs to thermodynamics. Both because of its own inherent importance and because of the insight which the discussion of it will afford into the nature of physical concepts generally, it will be worth our while to develop the idea of temperature in some detail.

As a qualitative definition, we may say that temperature is a name for the degree of the hotness or coldness which is a directly sensed characteristic of every physical object with which we have contact. To go beyond this, however, and provide the basis for a quantitative definition, it is necessary to consider the phenomenon of thermal equilibrium. The concept of thermal equilibrium applies only to systems which do not interact chemically or exchange matter. Experience shows that when two

such systems, initially at different temperatures or degrees of hotness—we have as yet no quantitative measure of temperature—are brought into contact, then in the course of time they settle down to a state of equilibrium at a common temperature. Provided that no effects are produced in the surroundings, the equilibrium temperature is always intermediate between the two initial temperatures. This fact is of great importance, for it means that, if each of two systems is in thermal equilibrium with a third, then they are in thermal equilibrium with one another: the common temperature which they have before being brought together must be the same as their equilibrium temperature. This proposition, obvious as it seems, is by no means a physical necessity; for example, no such condition prevails in the case of electrical equilibrium involving metals and electrolytes, but, if it were not true, it would be wholly impossible to give any definition of the temperature of an object.

The discussion of thermal equilibrium in terms of our sensations of hotness, though it affords a justification for the concept of temperature, still does not of itself tell us how to measure it. To provide for a quantitative definition of temperature, we must introduce properties which are themselves defined in terms of quantitative operations, properties such as pressure, volume, surface, composition, and electrical resistance. This may be done on the basis of the observation that the temperature (degree of hotness) of any system is correlated with the values of certain other variables of the kind just referred to, i.e., variables subject to quantitative measurement. This makes it possible to define the temperature of the system quantitatively in terms of these variables. For example, in the case of gases the temperature appears to depend solely on pressure and volume, and at constant pressure the volume increases continuously with temperature. The volume of a gas at any chosen standard pressure may, therefore, be used as a measure of its temperature; e.g., the temperature may be taken as equal or proportional to the volume. Once the temperature of any one system under a restricted set of conditions has been defined in some such way as this, then, in accordance with the principle of thermal equilibrium, the system may be used as a thermometer to measure the temperature of any other system. It is then possible, on the basis of experiment, to determine the mathematical relation between the temperature of the new system and the other variables characterizing it. The relation is called the equation of state of the system. When once the equation of state of a system is determined, then it is possible to specify the temperature of the system from a knowledge of the other variables, without recourse to our thermometer, and the system may itself be used as an alternative thermometer. We may, in particular, use it to determine the complete equation of state of the original thermometer

employed to define temperature in the first place. When we proceed in this way, no inconsistency is ever encountered, and there is complete justification of the principle of thermal equilibrium, which was based on our qualitative estimations of temperature. The attainment of thermal equilibrium might, of course, be thought of wholly in terms of variables other than temperature, and in this way it would be possible, as has been done by Caratheodory, to define temperature without any reference whatever to our sensations of hotness. In respect to conciseness and elegance, this procedure is probably preferable to that which we have adopted.

Temperature as we have just defined it is called empirical temperature, to distinguish it from absolute temperature, a related quantity which has its origin in the Second Law of Thermodynamics, as will be explained later. There is a considerable amount of latitude in the definition of empirical temperature. For example, if we choose a liquid, like mercury, for our standard thermometer, and are content to regard temperature as dependent only on pressure and volume, then we may define temperature in terms of *any* convenient function of pressure at some constant volume, or of volume at some constant pressure. Similarly, if we use the electrical conductivity of a piece of platinum as our standard of temperature, we may define temperature in terms of any suitable function of conductivity at either constant pressure or constant volume. In accordance with our chosen definition, the equations of state of mercury, platinum, and all other systems will be different. There is another element of arbitrariness in the definition of temperature. The number of variables which it is necessary to take into account will depend on the degree of accuracy in which we are interested. Thus if we use mercury for our standard thermometer, it will not suffice for the very highest precision to define temperature in terms of volume at constant pressure; if we attempt to do so, we shall find that it is impossible to formulate an exactly obeyed equation of state for any system. This is because our definition fails to include surface. Actually the space occupied by a molecule at the surface of a liquid is different from that occupied by one in the interior; consequently the volume of the mercury is, to a slight extent, dependent on the surface volume ratio, as well as on pressure and temperature. One of the reasons for using gas thermometers is that, in this case, the only variables which need to be considered are pressure and temperature, although, for the highest accuracy, we cannot assume Boyle's law to hold exactly.

Dimensions of Temperature

At this point the question presents itself as to the dimensions of temperature. Without going into the theory of dimensions, we may point out

that our choice of fundamental units is, to a large extent, arbitrary and is limited only by the requirement that our equations balance dimensionally, a requirement introduced into physics for reasons of convenience, but in no sense constituting a logical necessity. The dimensions of any quantity are determined by the equations which define it in terms of the fundamental units. It is always possible to preserve the dimensional balance of our equations by increasing the number of fundamental units (dimensions) and, at the same time, introducing compensating dimensional constants. This is well illustrated in the case of temperature. Suppose that we define temperature in terms of a gas which obeys Boyle's law. The equation of state of the gas will depend on the precise nature of our definition, but, owing to the fact that Boyle's law prevails, it will always be of the form

$$\theta = f(PV)$$

where θ signifies temperature, P pressure, V volume, and f an arbitrary function. If we set $f(PV) = PV$, then θ has the dimensions of work or energy. If we set $f(PV) = (PV)^2$, then it has the dimensions of energy squared. If we set $f(PV) = (1/K)PV$, then we have the option of making K a numerical constant, independent of the choice of the units of P and V , in which case θ again has the dimensions of work; or we may make θ a new fundamental unit, independent of the fundamental mechanical units defining P and V , in which case K becomes a dimensional constant having the dimensions of energy per unit of temperature, i.e., energy per degree; or we may give θ some arbitrary definition in terms of any of the fundamental units, in which case K is again a compensating dimensional constant. For example, odd as it sounds, we might give θ the dimensions of time, in which case K would necessarily have the dimensions of energy per unit time, or power. From a logical point of view, all these definitions are equally valid. The only basis for choosing among them is convenience. So we see that it is meaningless to ask, "What are the dimensions of temperature?" We can only ask, "What convention is actually adopted as to the dimensions of temperature?" The answer to this question is that, however it is defined and whatever be the choice of our standard thermometer, temperature is always introduced as a new fundamental unit, accompanied by a corresponding dimensional constant. If we define temperature as being proportional to PV for an ideal gas, then this constant is known as the gas constant, and has the dimensions of energy per degree. Owing to the fact that temperature defined in this way turns out to be the same as the absolute temperature, ideal gases are of great interest and the gas constant plays a large role in physics.

It will be seen from the preceding discussion what a complicated and subtle thing the familiar concept of temperature becomes when we try

to give it a precise definition. Its apparent simplicity is to be attributed largely to the frequency with which we use it in everyday life and the homeliness of the ordinary thermometers with which we measure it. But simple instruments are not always associated with simple concepts. The same experience which we have encountered in attempting to give a rigorous definition of temperature also arises in the case of many other familiar and apparently simple physical concepts when we try to give them an exact formulation.

First Law

The First Law of Thermodynamics is familiar to every student of elementary physics as the "principle of the conservation of energy." Familiar as it is, however, its full significance is not always understood as well as it might be, and many students probably have an unfounded sense of confidence in their understanding of it. The principle is often stated by saying that, when any closed system passes from an initial state to a final state, the sum of the heat absorbed and the work done is fixed. This implies that in an isolated system, which neither absorbs heat nor does work, there is some function of the state of the system which remains constant. This is called its energy. The principle finds its simplest illustration in the case of an isolated system in which only mechanical effects are involved. This is exemplified by a pendulum swinging in a vacuum from a frictionless suspension. In this case we know that the sum of the potential and kinetic energies of the bob remains constant. At the top of the swing, where the pendulum is, for an instant, at rest, the potential energy is at a maximum and the kinetic energy is zero; at the bottom of the swing, where the bob is moving fastest, the potential energy is at its minimum and the kinetic energy at its maximum. Of the two forms of energy either one always increases at the expense of the other. It is shown in discussions of mechanics that this state of affairs in which there is a constancy of the two forms of energy always prevails in isolated mechanical systems where the forces acting between the particles are of the type illustrated by gravitational and electrostatic attractions, that is, central forces which mathematically may be derived from a potential. If all the phenomena of the universe could be shown to be ultimately mechanical in character, just as heat can be interpreted in terms of the motions of the molecules, all the forces involved being of the type in question, then the law of the conservation of energy would follow at once from the principles of mechanics, and we could state without further consideration that in any isolated system the total amount of energy remains constant and that the energy of a closed system can only vary as a result of exchanges of energy with the surroundings.

Thermodynamics, however, proceeds without any such special interpretation. Indeed, much of its beauty and power arises from the fact that it makes use of a minimum number of physical hypotheses and models, such as the mechanical theory of heat. From this point of view, in the absence of any independent definition of the concept of heat, it will be seen that the First Law, as we have stated it, is no law at all, but simply a definition. It embodies no information based on experience, since it involves only one experimentally determinable quantity, namely work. Actually, reflection shows that we can develop the First Law without introducing the idea of heat at all, simply on the basis of facts relating to the work done under different conditions when a system passes from one state to another. The law may then be given the following statement. *It is always possible to effect the transition of a system from one state to another, provided the direction of the transition is not specified, solely by the performance of mechanical work, without leaving behind any change in the state of any other system; and when this is done, the amount of work involved is independent of the way in which the work is performed.* Suppose, for example, we are required to raise the temperature of a certain mass of water by a certain amount, the pressure being held constant. This may be done in either one of the following three ways (among others): (1) We may stir the water with a paddle as in Joule's experiment on the mechanical equivalent of heat. The mass of the paddle can be made indefinitely small in comparison with that of the water, so that the change of temperature of the paddle may be neglected, the only significant effects being the heating of the water and the performance of work. (2) We may expend the work in turning an electric generator which supplies a current to heat the water. Here again the mass of the resistor which carries the current can be made negligible. (3) We may expend the work in compressing, and thereby heating, a gas in an insulated chamber, having previously permitted the gas to expand into a vacuum. The heated gas may then be allowed to come to thermal equilibrium with the water. The conditions may be so chosen that at the end of the process the water has the desired temperature and the gas its original temperature and pressure. The First Law tells us that, as an experimental fact, the mechanical work involved is the same in each case. Of course, it is a matter of experience to determine what variables it is necessary to consider in defining the state of the system, in order that the First Law may be satisfied. These need not be identical with those which enter into the equation of state of the system; for example, in the case of an ideal gas, temperature alone suffices; in the case of a metal condenser carrying a fixed electric charge, temperature, pressure, and the spacing of the plates are all involved. The important fact is that if we take account of an adequate number of

experimentally measurable variables, the law is satisfied. This is an experimental result which could not have been predicted beforehand. It leads to the proposition, which is sometimes used as a statement of the First Law, that it is impossible to construct an engine which will operate in a cycle (i.e., returning to the state from which it started), and give rise to no other result than the performance of mechanical work. Such a hypothetical machine is called a perpetual motion machine of the first class.

The amount of work done *by* the system in passing from one state to another, when no effects remain in other systems, may, on the basis of the First Law, be written as

$$W = E_1 - E_2 \quad (1)$$

where E is a function of variables which define the state of the system, the subscripts 1 and 2 being used to denote the values of the function for the initial and final states of the system, respectively. The function E is called the energy of the system.

It will be seen, from equation (1), that we can only determine differences of energy; that is, we can specify the energy of a system only in relation to some particular chosen reference state, or ground state; in other words, energy contains an additive constant not accessible to experiment. Another point about energy which should be emphasized is that it is an extensive property: if two identical systems undergo identical changes of state, under the conditions we have assumed (no changes to be left in other systems) the work involved will be the same for both; that is, if we double the mass of a system, we double the work required to effect a given change of state. Consequently, we double the energy.

Under other conditions than those assumed above, when, namely, the transition is associated with changes in the state of other systems, equation (1) is in general not fulfilled; that is, the work done is not equal to the change in energy of the system in question. The difference is then by definition the heat absorbed, Q :

$$Q = E_2 - E_1 + W \quad (2)$$

It will be seen that equation (2) is the algebraic form of our original statement of the First Law, which was pronounced unsatisfactory, as amounting simply to a definition of heat. Although, in accordance with our development of the subject, equation (2) still remains simply a definition of heat, it should not be overlooked that it embodies the First Law in so far as it makes use of the concept of energy, which originates from that law in accordance with equation (1). Consequently, when we make

use of this equation we are, in fact, introducing the First Law as an experimental principle. Equation (2) is often written in differential form as

$$dQ = dE + dW \quad (3)$$

which, for many purposes, is more useful. It is significant that equation (2) provides a definition of heat which is wholly independent of the concept of temperature, for the state of the system, which determines E , may be specified entirely in terms of other variables, such as pressure and volume.

Equation (2) shows that the dimensions of heat are those of energy. Instead of expressing heat in terms of ergs, however, it is often convenient to express it in terms of a larger unit known as the calorie. This is the amount of work necessary to raise the temperature of 1 gram of water 1°C and is equal to 4.183×10^7 ergs.¹ When the temperature of 1 gram of water rises 1°C as a result of the attainment of thermal equilibrium with some other body, no mechanical work being done on the water, we say that the water has received an amount of heat equal to 1 calorie. This suggests another way of looking at the subject, which is, in fact, the one that was first adopted historically. We might have developed the concept of heat independently of the First Law on the basis of changes of state of some standard system in the absence of mechanical work. Then we could have stated the First Law according to the original proposal, corresponding to equation (2). On the whole, however, the procedure would have been more difficult and confusing than that we have adopted.

It is often possible, on the basis of the principles of mechanics or of electricity and magnetism, to express the work done by a system in terms of variables describing the state of the system. This will generally be true, and will only be true, if the system is in equilibrium with its surroundings as it undergoes a change of state. Under these circumstances, the change of state is said to be a "quasi-static" process. Thus, if the system is a compressible fluid which undergoes an infinitesimal quasi-static change of volume, dV , then the work done is given by $P dV$, where P is the pressure in the fluid. If the system is a rod or spring which undergoes a quasi-static change of length, dL , the work done is $-T dL$, where T is the tension of the spring. If the system is a liquid whose surface is changed quasi-statically by an amount dS , then the work done is $-\gamma dS$, where γ is the surface tension of the liquid. Or, if the system is a battery which delivers a charge, dq , quasi-statically, the work done is $E dq$, E being the voltage of the battery. The reason for specifying that the

¹ For the highest accuracy, we must, of course, specify the initial temperature of the water and the pressure.

change of state be quasi-statical if the work is to be expressible in the manner just illustrated will be clarified by considering the case of a compressible fluid. If this expands without being in equilibrium with its surroundings, the external pressure, which determines the amount of external work done, will be less than the internal pressure, P , and the work done will be less than $P dV$. Moreover, the process will involve turbulence and inequalities of pressure in the fluid so that it will be impossible to assign a precise value to P . In the limiting case where the expansion takes place into a vacuum, no work at all will be done, although the internal pressure will never vanish at any point in the fluid. A quasi-statical process, consisting as it does of a succession of states in which the system is always in equilibrium with the external restraining forces, represents an ideal process which cannot occur in any finite time and has no preferred direction. It is, therefore, reversible, and after it has occurred, first in one direction and then in the opposite direction, everything will have returned to its original condition. It is to be thought of as a limiting process approached by either one of two opposite processes as the external forces approach the equilibrium values from opposite sides, the speed tending to zero.

Whenever a system undergoes a change of state during which there is no exchange of heat with the surroundings, that is, when the only external effect is the performance of mechanical work, the process is said to be *adiabatic*. In accordance with equation (3), therefore, for every element of an adiabatic process

$$dE + dW = 0 \quad (4)$$

This equation is the embodiment of the First Law as it applies to adiabatic processes. Such processes are particularly important in discussions of the Second Law of Thermodynamics, to which we now turn.

Second Law

A clue to the Second Law is to be found in the restriction regarding direction which we introduced into our statement of the First Law. In the discussion of the First Law, we outlined three different ways in which it would be possible to raise the temperature of a body of water solely by the performance of mechanical work, without leaving any changes in the state of any other system. Had we been required to lower the temperature of the water instead of raise it, it would have been impossible to effect the task in this way. This represents the fact that the generation of heat by friction, the production of heat by an electric current flowing through a resistor, and the free expansion of a gas are all irreversible processes. That is to say, they are processes which, once they have occurred, can never be undone if we take account of all the effects involved. The Second Law is concerned with the phenomenon of irreversibility.

The Second Law may be stated in a number of ways. The formulation of it given by Planck is as follows: It is impossible to construct an engine which will work in a cycle (thereby returning to its original state) and leave behind as a result of its operation no effects except the cooling of a heat reservoir and the performance of mechanical work. Such an engine has been called a perpetual motion machine of the second class, to distinguish it from a perpetual motion machine of the first class, which is forbidden by the First Law. By ruling out a perpetual motion machine of the second class we are in effect saying that it is impossible to convert heat completely into work without producing any other effects. The second law as stated by Planck leads us at once to pronounce as irreversible certain processes which in the light of experience we recognize to be such: e.g., the generation of heat by friction and the free expansion of a gas. Consideration shows that if we had any means of reversing either one of these processes without leaving behind any other effects, such means would constitute a perpetual motion machine of the second class.

Another, somewhat more general and abstract, statement of the Second Law is that due to Caratheodory: Arbitrarily close to any given state of any closed system there exists an unlimited number of other states which it is impossible to reach from the given state as a result of any adiabatic process, whether reversible or not. This statement is to be interpreted in terms of the geometrical representation of the state of a system by a point in an n -dimensional space, the position of the point being fixed by n Cartesian coordinates, one for each of the n *independent* variables required to define the state of the system. (In the case of a gas, the number of such variables is two, e.g., pressure and temperature, and the state of the gas can be represented by a point in a plane; but in general the value of n will be greater than this, and it will be necessary to make use of a space of a greater number of dimensions to represent the state of a system.) The law means, then, that in any arbitrarily chosen element of volume in this space, which contains the given point, there are infinitely many other points which it is impossible to reach, starting from the given point, by any adiabatic process.

From this statement of the law as it stands, it is not immediately possible to say just what states are forbidden. On the basis of experience, however, we see how the principle is illustrated in certain familiar cases. If we consider a mass of any substance at pressure P and temperature θ , we recognize at once as an inaccessible point one having the coordinates P and θ' , where $\theta' < \theta$, or if we consider a mass of a nearly ideal gas such as hydrogen at V and θ , then we recognize as an inaccessible point one having the coordinates V' and θ , where $V' < V$. If these points were *not inaccessible*, we should be able to reverse the generation of heat by friction and the free expansion of a gas.

From either of the above statements of the Second Law, that of Planck or that of Caratheodory, it may be deduced that for any system there exists a function, S , of the state of the system which has the property that, when the system undergoes any process whatever, then for each element of the process

$$dS \geq \frac{dQ}{T(\theta)} \quad (5)$$

where Q refers to the heat absorbed and $T(\theta)$ is a single valued function of the empirical temperature, θ . The term $T(\theta)$ is what is known as the absolute temperature, and S the entropy. The inequality sign corresponds to a spontaneous and, therefore, irreversible process, and the equality sign to the limiting case of a reversible process. Equation (5) is the mathematical expression of the Second Law in the same way that (3) is the mathematical expression of the First Law. Just as the First Law leads to the concept of energy as an attribute of a system determined uniquely by its state, so the Second Law leads to the concept of entropy as another attribute of a system determined uniquely by its state. Both entropy and energy, as quantities determined solely by the state of a system, are sometimes spoken of as potentials. Expressed as functions of the variables defining the state of the system they both satisfy the conditions for an exact differential.² There is, however, a feature of entropy represented by the inequality sign in equation (5) which has no counterpart in the case of energy. This corresponds to the fact that entropy is associated with irreversibility and direction and that the Second Law is in a sense a historical principle, whereas the First Law embodies simply a principle of conservation. In an isolated system the entropy is capable of increasing spontaneously out of nothing whereas the energy necessarily remains constant.

We shall now proceed to develop the concepts of entropy and absolute temperature and deduce the fundamental equation (5), as the mathematical embodiment of the Second Law, from the statement of the law as given by Caratheodory.

We shall begin by considering the states accessible from a given state, P , of a system by any reversible adiabatic process. It will be clear that these states, together with P , will be represented by a continuous set of points in an n -dimensional space having one dimension for each of the n variables $x_1, x_2 \dots x_n$ necessary to fix the state of the system. Now

² If y is regarded as a function of x_1 and x_2 , then the condition that dy is an exact differential is that

$$\frac{\partial}{\partial x_1} \left(\frac{\partial y}{\partial x_2} \right) = \frac{\partial}{\partial x_2} \left(\frac{\partial y}{\partial x_1} \right)$$

these points must occupy either a volume or an $(n - 1)$ -dimensional surface in the n -dimensional space containing P . No further limitation of their distribution need be considered, since only one restriction among the variables can arise from the single condition of a reversible adiabatic process.* If the points occupied a volume, containing P as an interior point, then there would certainly be a violation of the Second Law (as given by Caratheodory). On the other hand, if P were a point in the bounding surface of the volume, then, though there would still be points inaccessible from P , the law would be violated for any point situated *within* the volume, since if two points, P' and P'' , are both accessible from P by a reversible adiabatic process, either one is accessible from the other by way of P . It follows, therefore, that the points accessible from P must all lie in a surface passing through P . Such a surface is called an adiabatic surface. (If $n = 2$, as in the case of a gas, this surface is of course a line.) It may be extended indefinitely in the n -dimensional space. Any curve in the surface corresponds to a reversible adiabatic process, and conversely any adiabatic process involving a point situated in the surface corresponds to such a curve. The adiabatic surface will be defined by a relation between the n variables, which we may write:

$$f(x_1, x_2 \cdots x_n) = \text{constant} \quad (6)$$

Consequently any curve corresponding to a reversible adiabatic process will conform to the relation

$$\frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 + \cdots + \frac{\partial f}{\partial x_n} dx_n = 0 \quad (7)$$

But it must also satisfy the condition for an adiabatic process. By the First Law, this condition is

$$dQ = dE + dW = 0 \quad (8)$$

Since we are concerned only with reversible processes, we may express dW as well as dE as a differential expression in terms of the variables $x_1 \cdots x_n$ which define the state of the system. If, having done this, we compare the two equations (7) and (8), it is clear that the coefficients of the infinitesimals in the two cases must be proportional to one another.

* It might be thought that this condition, involving a relation between the variables, would automatically imply that the points lie on an $(n - 1)$ dimensional surface. This is not so, since the condition is expressed in terms of a differential equation and it is by no means clear whether this can be integrated. Indeed it is precisely with this fundamental point that we are here concerned.

The proportionality factor, which we may call M , will be a function of $x_1, x_2 \dots x_n$. Consequently we can write

$$df = \frac{dE + dW}{M} \quad (9)$$

or

$$f = \int \frac{dE + dW}{M}$$

The factor M is, therefore, an integrating factor (sometimes called an integrating denominator) for the expression

$$dE + dW = dQ$$

and we see that such a factor must always exist in the case of any reversible process. The establishment of this fundamental point marks the first step in the derivation of equation (5).

At this point, before going on, it is relevant to call attention to the fact that the proposition we have just reached regarding the existence of an integrating factor would hold in any case if we were concerned with a system characterized by only two independent variables, for it is a general mathematical principle that any differential expression in two variables has an integrating factor. It is only when we consider systems the definition of whose state requires more than two variables (systems having, as we may say, more than two degrees of freedom) that the principle becomes an expression of the Second Law.

This explains a fact that would otherwise remain something of a puzzle, namely, that it is possible, on the basis of the First Law alone, to derive an expression for the efficiency of a Carnot engine consisting of a perfect gas (with only two degrees of freedom), though in the general case, where the nature of the engine is not specified, this can only be done with the aid of the Second Law. Such a derivation is often found in books on thermodynamics prior to the discussion of the Second Law.

The next step in the development of the Second Law is to introduce the concepts of absolute temperature and entropy. In doing this, we begin by calling attention to another general proposition regarding integrating factors. This is the proposition that, if an expression admits of one integrating factor, M , then it admits of an infinity of others also, each corresponding to a different integral, f . This is illustrated by the expression

$$dx + \frac{x}{y} dy + \frac{x}{z} dz$$

for which it is easy to verify that either $M = 1/yz$, corresponding to $f = xyz$, or $M = x$, corresponding to $f = \ln(1/xyz)$, is an integrating factor. There are infinitely many others. The various integrating factors

are, however, not independent of one another, owing to the fact that the corresponding integrals are functionally related. Thus if f' and f'' are two such integrals, f' is always expressible as a function of f'' , and vice versa. If M' and M'' are the corresponding integrating factors, we can write

$$M' df' = M'' df'' \quad (10)$$

or

$$M' = M'' \left(\frac{df''}{df'} \right)$$

Since df''/df' may be expressed as a function of either f' or f'' , this shows that, although in general M' and M'' will both be functions of all the variables x_1, x_2, \dots, x_n , M'/M'' is expressible as a function of f' or f'' only. As between the various integrating factors of which an expression is susceptible, there is in general nothing to choose except on grounds of simplicity. In the case of thermodynamic systems, however, owing to the phenomenon of thermal equilibrium, there is one particular integrating factor for the fundamental expression $dE + dW = dQ$, formulated for a reversible process, which is of unique significance because it is the same for all systems and depends only on the empirical temperature. This is what appears in equation (5) as the absolute temperature.

The existence of absolute temperature, associated with entropy, as an integrating factor for $dQ = dE + dW$ for a reversible process can be established in the following manner. We begin by considering a system composed of two bodies in thermal equilibrium with one another. We assume the state of each body to be definable by two variables, which may be conveniently chosen as the empirical temperature, θ , and the volume, V ; θ will of course be the same for both bodies. For each of the bodies considered separately, and for the whole system, there will exist integrating factors. Consequently, we may write

$$\begin{aligned} M_1 df_1 &= dE_1 + dW_1 \\ M_2 df_2 &= dE_2 + dW_2 \\ M df &= dE + dW \end{aligned} \quad (11)$$

these three equations referring to body 1, body 2, and the whole system, respectively. Now, by the First Law,³

$$dE_1 + dE_2 + dW_1 + dW_2 = dE + dW \quad (12)$$

Therefore, df , which we know to be exact, is given by:

$$df = \frac{M_1}{M} df_1 + \frac{M_2}{M} df_2 \quad (13)$$

³ Since in the reversible process any work done by the first body on the second must be equal and opposite to that done by the second on the first, $dW = dW_1 + dW_2$. Also, in the absence of any mutual energy of the two bodies, $dE = dE_1 + dE_2$.

This shows that f is a function of f_1 and f_2 only, so that M_1/M and M_2/M must be functions of these two variables exclusively; M_1 , on the other hand, will depend only on θ and V_1 or, since f_1 is a function of θ and V_1 only, on θ and f_1 . Similarly, M_2 will depend only on θ and f_2 ; M , however, will depend on θ and both V_1 and V_2 , or on θ , f_1 , and f_2 . Consideration shows that these conditions of dependency can be satisfied if and only if M_1 , M_2 , and M are expressible in the following way:

$$\begin{aligned} M_1 &= cT(\theta) \frac{\psi_1(f_1)}{c} \\ M_2 &= cT(\theta) \frac{\psi_2(f_2)}{c} \\ M &= cT(\theta) \frac{\psi(f_1, f_2)}{c} \end{aligned} \quad (13.1)$$

where c is an arbitrary constant. If therefore we introduce as a definition

$$S_1 = \int \frac{\psi_1(f_1)}{c} df_1 \quad (14)$$

we can write for body 1, on the basis of (11),

$$M_1 df_1 = T dS_1 = dE_1 + dW_1 \quad (15)$$

If we identify T , which is a function solely of the empirical temperature and must be independent of the particular properties of either system, with absolute temperature, and S_1 with the entropy of body 1, we see that this equation represents, for body 1, just what we set out to establish. Exactly the same considerations apply to body 2. As regards the whole system, it follows from equation (12) that

$$\frac{dE + dW}{T} = dS_1 + dS_2 \quad (16)$$

If we introduce $S = S_1 + S_2$ as the entropy of the whole system, this becomes identical with the expression for either of the component bodies; S as the sum of S_1 and S_2 is of course completely determined by the state of the system. The principle, therefore, holds for the whole system, as well as for each component body.

The same result may also be shown to apply to any system whatever, provided only we take account of an adequate number of variables. If, therefore, we make use of the concept of heat, which results from the First

Law, the Second Law tells us that, in the case of *any* system, there exists a function, S , called entropy, of the variables which specify the state of the system, such that

$$dS = \frac{dQ}{T} \quad (17)$$

where dQ is the heat absorbed during any *reversible* process and T is a universal function of the empirical temperature, called absolute temperature. Equation (17) is the same as equation (5), with the inequality sign omitted. We have succeeded, therefore, in our derivation of that equation in so far as it applies to reversible processes.

At this stage of the argument, it is worth pausing to discuss a point which might otherwise be a source of some perplexity to the reader. We have pointed out that, if an expression admits of one integrating factor, it admits of an infinity of others also. This must, of course, apply to dQ . Why then should there not be a variety of integrating factors for dQ , each, to be sure, of the form given by equation (13), but each involving different functions, $T(\theta)$ and $\psi(f)$? If this were so, then the uniqueness of absolute temperature (and entropy) would disappear. The answer to this is that there is indeed an unlimited number of integrating factors for dQ , each of the form $T(\theta)\psi(f)$, and each corresponding to a different integral, f , but that the function $T(\theta)$ involved in each is exactly the same. The differences reside wholly in the $\psi(f)$. A simple proof of this is as follows. Suppose that there are two such integrating factors, M' and M'' , corresponding to the integrals f' and f'' , respectively, and that these involve different temperature functions, $T'(\theta)$ and $T''(\theta)$. Then

$$M' = T'(\theta)\psi'(f') \quad \text{and} \quad M'' = T''(\theta)\psi''(f'')$$

But by (10)

$$M'' = M' \frac{df'}{df''}$$

where df'/df'' may be expressed as a function of either f' or f'' . Consequently

$$T''(\theta)\psi''(f'') = T'(\theta)\psi'(f') \frac{df'}{df''}$$

Since the functions df'/df'' and $\psi'(f')$ may both be expressed wholly as functions of f'' , this demands that $T''(\theta) = T'(\theta)$ and that $\psi''(f'')/\psi'(f') = df'/df''$. So we see that the two temperature functions are identical. It is also instructive to verify that, although the integrals f' and f'' are different, the corresponding entropies, defined as

$$S' = \int \psi'(f') df' \quad \text{and} \quad S'' = \int \psi''(f'') df''$$

are identical like the temperature functions. This follows immediately when we write

$$\begin{aligned} S'' &= \int \psi''(f'') \frac{df''}{df'} df' = \int \psi''(f'') \frac{\psi'(f')}{\psi''(f'')} df' \\ &= \int \psi'(f') df' = S' \end{aligned}$$

It is apparent, therefore, that absolute temperature and entropy are both unique functions.

The question might also arise whether we could not have employed the same line of argument as that already given to the case where the systems are at pressure equilibrium in order to establish the existence of an integrating factor depending only on the pressure. The answer is that we could, but actually equilibrium of two systems does not demand equality of pressure as it does equality of temperature; therefore, the procedure does not lead to any universal function and has no significance.

The task of determining the absolute temperature as a function of the empirical temperature, θ , will be accomplished if we can discover an integrating factor, depending only on θ , for *any one* particular system. A perfect gas serves the purpose. A perfect gas is a hypothetical substance whose properties are assigned on the basis of an extrapolation from the properties of actual gases as the temperature is increased and the pressure is reduced. It is a substance which obeys Boyle's law, and whose energy depends only on the temperature.⁴ For such a substance, undergoing a reversible change of volume,

$$TdS = dE + dW = C_v d\theta + P dV \quad (18)$$

in which C_v is a function of θ only. Since, by Boyle's law, PV is also a function of θ only, say $t(\theta)$, we have

$$dE + dW = C_v d\theta + \frac{t(\theta)dV}{V} \quad (19)$$

It will be seen that $t(\theta)$ is an integrating factor of this expression. It is, therefore, proportional to the absolute temperature. Consequently, if we define the empirical temperature as being proportional to PV for a perfect gas, then it becomes proportional to the absolute temperature. The proportionality factor simply has the effect of fixing the size of the degree. In this connection the reader should refer to our earlier discussion of the concept of temperature. Under ordinary conditions, such gases as hydrogen and nitrogen conform to the behavior of perfect gases very closely, and may, therefore, be used to define the absolute temperature scale with a high degree of approximation in accordance with the equation $PV = RT$. At low temperatures, however, all gases show significant departures from ideality, and in order to fix the scale of absolute temperature in the lower range it is necessary to study their behavior in detail. We need not go into this matter here. It should be pointed out, however, that the concept of absolute temperature points to the existence of an absolute zero, at approximately -273.1°C , absolute temperature being always positive.

⁴ Actually these two attributes of a perfect gas are not independent, but each is an expression of the other, as will be shown below in connection with equation (37.5).

With the establishment of a scale of absolute temperature, it becomes possible, by means of equation (17), to determine the entropy change of any closed system undergoing a reversible process on the basis of calorimetric measurements alone, without any knowledge of how to formulate $dE + dW$ in terms of variables describing the state of the system. It should be emphasized that this entropy change will depend only on the initial and final states of the system, being independent of the path. From such measurements it is possible therefore to assign the entropy of any system in any state with reference to some arbitrarily chosen standard state. Beyond this, however, in the absence of some further principle, we cannot go, entropy, like energy, containing always an additive constant inaccessible to measurement. As we shall see later, an additional principle, commonly called the Third Law of Thermodynamics, makes it possible to obtain absolute values of entropy, subject to certain limitations. This is intimately related to the statistical interpretation of entropy as representing the degree of randomness of a system.

It now only remains, in developing the Second Law, to consider entropy in relation to irreversible processes and to deduce the inequality sign in equation (5). In doing this we shall begin by restricting ourselves to adiabatic processes and show that in any irreversible adiabatic process the entropy of the system involved always increases. Consider a nearly perfect gas such as hydrogen at low pressure. Its state is completely determined by its entropy and volume. By suitably changing the external pressure, it will be possible to bring the gas to any desired volume, whether the conditions be adiabatic or not. Here we assume them to be adiabatic. Then, if the change of volume occurs reversibly, it follows from equation (17) that the entropy must remain constant. On the other hand, if it occurs irreversibly, the entropy must either always increase or always decrease, for if this were not so it would be possible to bring the gas to any desired volume and entropy, that is, to any desired state, as a result of purely adiabatic processes, which would contradict the Second Law. To determine which alternative holds, we have only to consider some one irreversible process. Now it can be seen from equation (18) that the free expansion of a nearly perfect gas into a vacuum, which occurs at constant temperature, involves an increase of entropy, for this equation will be obeyed with a high degree of approximation, and we are, of course, only concerned with a matter of sign. It follows, therefore, that whenever the gas undergoes any adiabatic change of state, its entropy must always increase, or, in the limiting case where the process occurs reversibly, remain constant. The same must be true of every other system. Otherwise, by combining such a system with a perfect gas, we could carry out processes equivalent to one in which the entropy of the gas

was made to decrease adiabatically. So we see that in any irreversible adiabatic process the entropy of the system involved must always increase. This shows that the inequality sign of equation (5) holds for all irreversible processes which are adiabatic.

In order to show that it holds for all irreversible processes generally, it is necessary to consider a process which is not adiabatic. Let us, therefore, fix our attention on a system (1) which undergoes an irreversible process during which heat is absorbed. We may suppose that the heat absorbed in this process is all derived from another system (2), which is at the same temperature as the first and undergoes only reversible changes of state. The two systems together constitute a larger system, which, since no heat is supposed to be exchanged with the surroundings, must conform to the considerations of the last paragraph. Consequently,

$$dS = dS_1 + dS_2 > 0$$

But, since system 2 undergoes only reversible changes of state while giving up the amount of heat dQ to system 1,

$$dS_2 = \frac{-dQ}{T}$$

Substitution of this expression for dS_2 in the preceding equation leads at once to the inequality $T dS_1 > dQ$, which for system 1 is identical with the inequality of equation (5). Since the argument is quite general, we may drop the subscript 1, and obtain (5) in its general form. In case the system in question is not at uniform temperature, which is possible since it is not supposed to be in equilibrium, we may rewrite the inequality (5) as:

$$T_{\max} dS > dQ$$

where T_{\max} is the maximum value of the temperature at any point in the system. If we add an equality sign to provide for the limiting case of a reversible process, this equation may be regarded as the complete embodiment of the Second Law:

$$T_{\max} dS \geq dQ \tag{20}$$

It has required some pages of rather close reasoning to arrive at the mathematical expression of the Second Law represented by equations (5) and (20). It might well be asked, why should we not have avoided all this trouble by stating the law by means of these two equations in the first place, instead of deriving them from a verbal principle. This would have been much easier. After all, is it not the primary aim of science to

arrive at equations which when developed yield results corresponding to the observations? Here the question arises as to whether there should be a unique relationship between the equations and the observations. In other words, is it necessary to show both that the observations imply the equations and that the equations predict the observations? If we answer this question in the affirmative, then there is some justification in proceeding as we have done. On the other hand, is it clear that the verbal statement from which we have deduced the equations is a logical consequence of our observations any more than the equations themselves? The answer is no, as must be the case for any such generalizations, verbal or mathematical, which are based on a limited number of experiments. It would seem, therefore, that we might have spared ourselves much pain by starting at once with the mathematical formulation of the Second (and, for that matter, the First) Law. Similarly, many books on electricity and magnetism begin directly with Maxwell's equations, and deduce other relations from them. The justification for proceeding as we have done is to be found not in the logic of the situation, but rather in the fact that by so doing we have certainly gained a deeper insight into the subject than would otherwise have been the case.

Irreversible Processes

The Second Law, as expressed in equation (5) or (20), enables us at once to establish the irreversibility of various processes. A good example is the flow of heat from a hotter body to a cooler body. Let us denote the two bodies by 1 and 2, and their temperatures by T_1 and T_2 , respectively, and suppose that an amount of heat, dQ , flows from 1 to 2, dQ being so small that the temperatures of the two bodies are not appreciably altered by the heat transfer. Let us further suppose that no other transfer of heat is involved and that any changes of volume which occur during the process are reversible. Then, by equation (17) for body 1,

$$dS_1 = - \frac{dQ}{T_1}$$

and for body 2

$$dS_2 = + \frac{dQ}{T_2}$$

For the system as a whole

$$dS = dS_1 + dS_2$$

Consequently since $T_1 > T_2$,

$$dS = dQ \left(\frac{1}{T_2} - \frac{1}{T_1} \right) > 0$$

Since, from the point of view of the system as a whole, the process is adiabatic, this means that it is irreversible. This result, of course, corresponds with one of the most universal facts of experience.

Another example is the process in which a weight falls to the bottom of a closed chamber. When the weight comes to rest, its kinetic energy is converted into heat, with the result that there is a rise of temperature of the system comprising the chamber and the weight. To show that this process is irreversible, we may compare it with an alternative one which leads to exactly the same result as far as the system is concerned. In this alternative process the weight, instead of falling freely, descends slowly (not necessarily reversibly) while raising an external opposing weight of nearly equal magnitude, and heat is absorbed by the system from the surroundings in an amount sufficient to produce the same rise of temperature as in the original process. Since the two processes lead to the same result as far as the system is concerned, the change in entropy of the system, is the same in both. For the second process, the integral

$$\int dS \geq \int \frac{dQ}{T_{\max}}$$

is greater than zero, since dQ and T_{\max} are both positive. In the first process, on the other hand, no heat is absorbed, so that

$$\int \frac{dQ}{T_{\max}} = 0$$

Consequently, for the first process

$$\int dS > \int \frac{dQ}{T_{\max}} = 0$$

and it follows from equation (20) that the process is irreversible, as we know it to be. The universal agreement between predictions of irreversibility based on equations (5) and (20) and the facts of experience may be regarded as the proof of the Second Law. One single case of disagreement would suffice to disprove the law, but none has ever arisen.

It should be emphasized that irreversible processes are such that, once they have occurred, it is impossible to restore things completely to their original condition. This does not mean, of course, that a given irreversible process, e.g., the free expansion of a gas, cannot be reversed at the expense of some other process involving other systems. After the reversal has occurred, however, there will remain changes in the state of other systems. If this were not so, then, for the larger system which includes these additional systems called into play, we should have a

decrease of entropy in the absence of any interaction with its surroundings, that is, as a result of an adiabatic process. This, however, is contrary to equation (5).

The steady increase of entropy in nature as a whole, accompanying the playing out of spontaneous and, therefore, irreversible processes, sets it off in marked contrast to energy, which remains unchanged in past and future alike. It makes it, in fact, as we indicated before, a kind of historical concept, related to the direction of time. In statistical mechanics, entropy is interpreted in terms of the probability or degree of randomness of the state of a system. The increase of entropy in nature associated with the passage of time means, therefore, an increase of randomness, and we may think of the course of history as representing a steady growth of disorder of the universe as a whole.

The intimate connection between the concepts of entropy, probability, and time were given a novel interpretation by A. S. Eddington. He suggested that our very recognition of the direction of time, that is, our distinction between past and future, is not something immediate but is derived from our apprehension of probability, related as it is to entropy; we recognize a later state of affairs in relation to an earlier one on the basis of the different degrees of randomness of the two. This point of view, however, appears to ignore the fact that it is only on the basis of a primitive sense of time's direction that we recognize a state of greater randomness as following one of less. Is it not possible to imagine a world in which there would be a steady increase, rather than a steady decrease, of order? This happens, however, not to be our world.

Maximum Work, Equilibrium, and Free Energy

The necessary and sufficient conditions for the reversibility or irreversibility of a process, which result from the Second Law, are contained in equation (20). After combination with (3), which is the expression of the First Law, this gives

$$T dS \geq dE + dW \quad (21)$$

assuming the system to be at uniform temperature. It is, of course, to be understood that the equality sign corresponds to reversibility, the inequality sign to irreversibility.

A significant fact which emerges from this relation is that the maximum work realizable from a process, when it occurs reversibly, is not equal to the decrease of total energy of the system involved, as might be supposed on the basis of the First Law, but is determined by the expression $(T dS - dE)$. Depending on whether dS is positive or negative, the maximum work will be either less or greater than the decrease of energy.

Any difference will, of course, be represented by heat given out or absorbed. This initially surprising result is a consequence of the Second Law. A familiar illustration of it is provided by the reversible expansion of a perfect gas which throughout the process remains in thermal equilibrium with a heat reservoir. As the gas expands, it does work ($W = \int P dV$), and, since the energy of the gas depends only on the temperature, which remains constant, this work must be entirely paid for by the heat absorbed from the heat reservoir. The gas, as it were, goes into debt to its surroundings to provide for the work of expansion. In other cases, such as various chemical reactions, where dS is negative, a certain amount of the energy liberated is unavoidably "spilled over" as heat, and is wholly unavailable for work even in a reversible process.

For certain purposes, it is convenient to break up the term dW into two terms, one, dW'' , representing the work done by the system against the pressure of the surroundings, and one, dW' , representing the work done against any additional forces that may be involved. Depending on whether or not any changes of volume which occur are reversible, dW'' will be equal to or less than $P dV$, where P is the pressure in the system. Consequently (13) may be written as

$$T dS \geq dE + P dV + dW' \quad (22)$$

with the same understanding as before regarding the equality and inequality signs. The term dW' may be regarded as the "useful" work done by the system over and above the unavoidable pressure volume work involved during the process. Just as relation (21) imposes an upper limit to the total work, so relation (22) imposes an upper limit to the "useful" work realizable from a process. This may be expressed in three different ways, one when it occurs at constant energy and volume, one when it occurs at constant entropy and volume, and one when it occurs at constant energy and entropy:

$$\begin{aligned} dW' &\leq T(dS)_{EV} \\ dW' &\leq -(dE)_{SV} \\ dW' &\leq -P(dV)_{SE} \end{aligned} \quad (23)$$

Relations (23) lead at once to three equivalent criteria of equilibrium for the important case of a system subject to no restraints except the pressure of the surroundings. In the absence of all restraints except pressure, dW' must necessarily be equal to zero, for there is nothing other than pressure for the system to work against. Consequently the three relations (23) formulated as equalities, corresponding to the case of reversibility in which the system is in a state of equilibrium, reduce to

$$(dS)_{EV} = 0, \quad (dE)_{SV} = 0, \quad (dV)_{SE} = 0 \quad (24)$$

Processes at constant energy and volume, or constant entropy and volume, or constant entropy and energy, such as those implied by equations (23), are not common and are hard to realize experimentally. Processes at constant temperature and pressure, on the other hand, corresponding to the case where the system involved is in pressure and temperature equilibrium with its surroundings, occur constantly and are easy to produce in the laboratory. Clearly, it is desirable to have an expression for the maximum "useful" work, over and above the pressure volume work, for such processes. This would also provide a useful criterion for equilibrium. Such an expression may be obtained very simply with the aid of a new variable first introduced into thermodynamics by Gibbs and now known as the *Gibbs free energy* or simply as the *free energy*. We shall denote it by F . It is defined as

$$F = E + PV - TS \quad (25)$$

Like the variables E , V , and S which it contains, F is an extensive property, completely defined by the state of the system. Unlike E and S , however, it contains not only an undetermined constant (from E), but also an undetermined constant (from S) multiplied by the absolute temperature. By introducing F into equation (21), we obtain at once

$$dF \leq -S dT + V dP - dW' \quad (26)$$

This equation is directly applicable to processes at constant temperature and pressure and shows that for such processes the "useful" work, dW' , done by the system is equal to or less than the decrease of free energy:

$$dW' \leq -(dF)_{T,P} \quad (27)$$

The appropriateness of the term *free energy* for the function F is apparent. Unlike the energy itself, the free energy is a quantity whose decrease sets a definite limit to the amount of "useful" work (i.e. the amount of work over and above the inevitable pressure-volume work) which the system is capable of doing when exposed to the temperature and pressure of the surroundings.

Like the relations (23), (27) leads to a criterion of equilibrium for a system free of all restraints except external pressure. For such a system dW' must be equal to zero, and consequently for a system in equilibrium

$$(dF)_{TP} = 0 \quad (28)$$

Equation (28) is, in every way, the equivalent of each of the three equations (24) as a condition of equilibrium. Owing to the fact that it is formulated in terms of constant temperature and pressure, however, it is more generally useful. It is easy to show that (28) holds as a criterion of

equilibrium even where the system consists of two or more phases, all at the same temperature, but each at a different pressure, provided only that we interpret the subscript P to mean constancy of pressure in each phase. This follows directly from equation (22), if we break up the terms dE , dS , and $P dV$ into a sum of terms, one for each phase: $dE = dE_1 + dE_2 + \dots$; $T dS = T(dS_1 + dS_2 + \dots)$;

$$P dV = P_1 dV_1 + P_2 dV_2 + \dots$$

Heat Content, or Enthalpy

In addition to the two fundamental thermodynamic functions, energy and entropy, it has proved useful to introduce another, namely, the free energy, dealt with in the last section. There is still one more function, the use of which considerably simplifies certain thermodynamic discussions. This is what is called by some authors enthalpy and by others heat content. We shall denote it by H . The heat content of a system is defined as

$$H = E + PV \quad (29)$$

It will be seen that it is an extensive property, and that, involving as it does the energy E , it contains an undetermined constant. In general, H is not a very useful function except for processes occurring at constant P ; but these are in practice the most important. Whenever a system undergoes a reversible change of state at constant pressure, $dH = dE + P dV$. If no other work is done than the pressure volume work, $\int P dV$, dH will be identical, by the First Law, with the heat absorbed. In the case of chemical reactions occurring at constant pressure, the change of heat content is the same as what is measured as the heat of the reaction. The reason for making use of the function H , rather than heat itself, is that the former is a thermodynamic function depending only on the state of the system, whereas the latter is not. On account of this, the heat content is susceptible to mathematical treatment not applicable to the heat itself.

Expressions for the Change of the Fundamental Thermodynamic Functions

As we have pointed out, the First Law ($dE = dQ - dW$) and the Second Law ($dS = dQ/T$ for a reversible process) provide a general basis for determining, respectively, the energy and the entropy of any system. In each case, however, the result will contain an undetermined constant corresponding to the arbitrary choice of a reference state. Subject to the limitations imposed by this fact, the values of energy and entropy so obtained may be used to fix the values of the enthalpy and free energy of the system in accordance with the fundamental definitions of these quantities, for the only other variables involved are P , T , and V . It is

not sufficient for our purposes, however, to leave the matter of the evaluation of the four thermodynamic functions at this very general level. On the contrary, it is important to derive expressions for various differential coefficients involving each of them, e.g., $(\partial S/\partial T)_P$ and $(\partial F/\partial P)_T$. Such expressions are not only useful in reckoning changes of the various functions, but are indispensable in the development of various basic thermodynamic relations. We shall devote this section to the task of developing a few of the more useful expressions and to discussing and applying them.

Let us begin with the free energy, F . This is completely determined by the state of the system. Since for a system of fixed composition we may assume the state to be defined by P and T ,⁵ we may write

$$dF = \left(\frac{\partial F}{\partial T}\right)_P dT + \left(\frac{\partial F}{\partial P}\right)_T dP$$

At the same time, if the system undergoes a reversible change of state without the performance of any work except pressure volume work, equation (26) becomes

$$dF = -S dT + V dP \quad (29.1)$$

Comparison of these equations shows that

$$\left(\frac{\partial F}{\partial T}\right)_P = -S \quad (30)$$

and

$$\left(\frac{\partial F}{\partial P}\right)_T = V \quad (30.1)$$

For some purposes, it is convenient to reformulate equation (30) in terms of H instead of S . From the definition of

$$F = E + PV - TS = H - TS$$

it follows that

$$-S = \frac{F - H}{T}$$

Consequently (30) may be rewritten as

$$\left(\frac{\partial F}{\partial T}\right)_P = \frac{F - H}{T}$$

If this equation is multiplied by $1/T$ and rearranged, it gives

$$\frac{1}{T} \left(\frac{\partial F}{\partial T}\right)_P - \frac{F}{T^2} = \frac{\partial}{\partial T} \left(\frac{F}{T}\right)_P = -\frac{H}{T^2} \quad (30.2)$$

⁵ If other variables are required to define the state of the system, these may be supposed to be held constant.

The terms $(\partial F/\partial T)_P$ and $(\partial F/\partial P)_T$ are the two most important first derivatives of F . Since volume is a directly measurable quantity, $(\partial F/\partial P)_T$ is subject to no uncertainty. On the other hand, $(\partial F/\partial T)_P$, being equal to $-S$, will contain an unknown constant unless we can establish a basis for determining absolute values of the entropy. This reflects the fact, to which we have already called attention, that the free energy itself contains not only an undetermined constant but an undetermined constant multiplied by the temperature. As we shall see later, the so-called Third Law of Thermodynamics makes it possible to obtain values of the entropy and consequently to determine this constant. Actually, the limitations ensuing from the presence of an undetermined constant in the entropy are not so serious as might be expected, since in most cases we are only interested in differences of free energy and entropy. This is brought out in the problem of deriving the so-called Clausius-Clapeyron relation, which may be accomplished with the aid of the two differential coefficients first obtained.

If a liquid and its vapor, or a solid and its vapor, or a solid and the corresponding liquid exist together at equilibrium, we know that the pressure is completely determined by the temperature. The question immediately arises, how does the equilibrium pressure vary with temperature. A direct solution of this problem results from the use of free energy. We shall distinguish the two coexisting phases by subscripts 1 and 2. Suppose that n units of the substance (it makes no difference whether the unit is the gram or the gram molecule) pass from phase 1 to phase 2 at P_0 and T_0 . Owing to the condition of equilibrium, the change of free energy associated with the process is zero:

$$\Delta F = nF_2 - nF_1 = 0$$

or

$$F_2 - F_1 = 0$$

Now the problem before us consists in finding a relation between P and T such that at any other pair of values which conform to the relation it will also be true that

$$\Delta F = 0$$

The condition for this is that

$$d\Delta F = 0 \quad \text{or} \quad dF_2 - dF_1 = 0$$

By writing

$$dF_2 = \left(\frac{\partial F_2}{\partial T}\right)_P dT + \left(\frac{\partial F_2}{\partial P}\right)_T dP = -S_2 dT + V_2 dP$$

and the corresponding expression for dF_1 , we see that this condition becomes

$$\frac{dP}{dT} = \frac{S_2 - S_1}{V_2 - V_1} \quad (31)$$

Here $S_2 - S_1$ is the difference between the entropy of a unit of substance in phase 2 and in phase 1, the two phases being at equilibrium. It is given by the heat absorbed divided by the temperature when a unit of substance passes (reversibly) from phase 1 to phase 2. At constant pressure, the heat absorbed will be equal to ΔH , so that $S_2 - S_1 = \Delta H/T$. Similarly, $V_2 - V_1$ represents the change of volume accompanying the process. Consequently, the solution of the problem becomes

$$\frac{dP}{dT} = \frac{\Delta H}{T \Delta V} \quad (32)$$

This fundamental relation is called the Clausius-Clapeyron equation. If we know ΔH and ΔV as functions of temperature, it may be integrated. It will be seen that the troublesome constants contained in S_1 and S_2 are not involved in the result, being without effect on the value of ΔH .

Let us now turn to differential coefficients involving the entropy. There are three of these which are of particular interest, namely,

$$\left(\frac{\partial S}{\partial T}\right)_P, \quad \left(\frac{\partial S}{\partial T}\right)_V, \quad \text{and} \quad \left(\frac{\partial S}{\partial P}\right)_T$$

Since in *any* reversible process, the entropy change is given by the heat absorbed divided by the absolute temperature, the first two of these may be at once equated to the specific heat at constant pressure and the specific heat at constant volume, respectively, each divided by the temperature:

$$\left(\frac{\partial S}{\partial T}\right)_P = \frac{C_p}{T} \quad \text{and} \quad \left(\frac{\partial S}{\partial T}\right)_V = \frac{C_v}{T} \quad (33)$$

As regards the third, since dF is an exact differential and consequently

$$\frac{\partial^2 F}{\partial T \partial P} = \frac{\partial^2 F}{\partial P \partial T}$$

it follows at once from (30) and (30.1), that

$$\left(\frac{\partial S}{\partial P}\right)_T = - \left(\frac{\partial V}{\partial T}\right)_P \quad (34)$$

The quantity $(\partial V/\partial T)_P$ is simply the change of volume with temperature at constant pressure, which can be easily determined by experiment or obtained from tables giving the coefficient of thermal expansion.⁶

⁶ This coefficient is usually given as fractional change of volume with temperature or pressure, i.e., as $\frac{1}{V} \left(\frac{\partial V}{\partial T}\right)_P$ rather than $\left(\frac{\partial V}{\partial T}\right)_P$.

A fourth differential coefficient, sometimes of use, may be immediately formulated in terms of measurable quantities on the basis of (34). This is

$$\left(\frac{\partial S}{\partial V}\right)_T = \left(\frac{\partial S}{\partial P}\right)_T \left(\frac{\partial P}{\partial V}\right)_T = - \left(\frac{\partial V}{\partial T}\right)_P \left(\frac{\partial P}{\partial V}\right)_T \quad (35)$$

Since by a mathematical principle applicable to any three quantities subject to a functional relationship,

$$\left(\frac{\partial V}{\partial T}\right)_P \left(\frac{\partial P}{\partial V}\right)_T \left(\frac{\partial T}{\partial P}\right)_V = -1$$

this may be alternatively and more compactly written as

$$\left(\frac{\partial S}{\partial V}\right)_T = \left(\frac{\partial P}{\partial T}\right)_V \quad (35.1)$$

Expression (34) is of particular interest. Since, in general, any substance expands with rise of temperature, equation (34) shows that increasing the pressure (at constant temperature) decreases the entropy. In terms of the statistical interpretation of entropy to be developed later, this means that compression must lead to a decrease of probability or an increase of order in the arrangement of the molecules; in the case of a liquid, there is an approach toward the regular packing characteristic of the crystalline state. There is one interesting and familiar exception to this, however. Water, in the temperature range 0° to 4° , shows an anomalous decrease of volume with increase of temperature. Since this implies an increase of entropy with pressure, it must mean that, in the temperature interval in question, pressure has the unexpected effect of breaking up an ordered arrangement of the molecules. An indication of why this should be so will be gained by consideration of ideas regarding the structure of water presented in Chapter 2.

Equation (34) has another interesting application in connection with the problem of relating the absolute temperature to the empirical temperature. As we have pointed out, even when we define empirical temperature in terms of a gas thermometer, it will show a significant divergence from absolute temperature at sufficiently low temperatures. The problem is, therefore, a very real one. If we introduce the empirical temperature, θ , we may rewrite equation (34) as

$$\left(\frac{\partial S}{\partial P}\right)_\theta = - \left(\frac{\partial V}{\partial \theta}\right)_P \left(\frac{d\theta}{dT}\right)$$

After multiplying both sides of this equation by T and rearranging, we obtain

$$\frac{1}{T} \frac{dT}{d\theta} = \frac{d \ln T}{d\theta} = - \frac{\left(\frac{\partial V}{\partial \theta} \right)_P}{T \left(\frac{\partial S}{\partial P} \right)_\theta} \quad (36)$$

The denominator of the right-hand member is, by the Second Law, the ratio of the heat absorbed to the change of pressure, at constant θ , in any reversible process. It can be measured in principle at least for the system constituting the thermometer as a function of θ . The numerator is, of course, an expression which will be given by the definition of θ . (For example, θ might be defined as proportional to the volume of the system at constant pressure, in which case $(\partial V / \partial \theta)_P$ would be equal to the proportionality constant.) We have, therefore, the means of integrating this equation on the basis of experimental data on the behavior of the system, and thereby determining T (or $\ln T$) in relation to θ .

A further illustration of the practical value of equation (34) is provided by the following problem. A mole of water at 100° is allowed to evaporate irreversibly into a region where the pressure is one-tenth of an atmosphere. What is the change of entropy? We may solve this problem by replacing the actual process by an alternative, reversible one leading to the same result. In this alternative process the water first evaporates reversibly at a pressure of one atmosphere. Then the pressure in the vapor is reduced, also reversibly, to one-tenth of an atmosphere. In the first part of the process, the change of entropy is given by the heat of vaporization of the water (9700 calories per mole) divided by the absolute temperature, 373° . In the second part of the process, the change of entropy on the basis of (34) may be set equal to $-\int_{P=1}^{P=1/10} \left(\frac{\partial V}{\partial T} \right)_P dP$. In order to evaluate this integral, we may assume the vapor to obey the ideal gas law:

$$PV = RT$$

with the gas constant $R = 1.99$ calories per degree. We then find

$$\left(\frac{\partial V}{\partial T} \right)_P = \frac{R}{P}$$

and

$$\int \left(\frac{\partial V}{\partial T} \right)_P dP = R \ln \left(\frac{P_2}{P_1} \right) = -R \ln 10$$

Making use of numerical values, we obtain for the net result of the whole process

$$\begin{aligned}\Delta S &= {}^{9700}_{373} + 1.99 \ln 10 \\ &= 26.0 + 4.6 = 30.6 \text{ calories per degree per mole}\end{aligned}$$

Let us now consider the differential coefficients involving the energy. By the same kind of reasoning as that involved in the derivation of (29.1) from (26), we obtain from (22) the equation

$$dE = T dS - P dV \quad (37)$$

which expresses E as a function of S and V . Equation (37) shows at once that

$$\left(\frac{\partial E}{\partial S}\right)_V = T \quad (37.1)$$

and

$$\left(\frac{\partial E}{\partial V}\right)_S = -P$$

In addition to these two fundamental expressions, it is desirable to derive others which show the way in which energy is related to the more readily controlled pairs of variables, T and P , and T and V . It is not difficult to obtain such expressions. As regards $(\partial E/\partial T)_V$, it follows at once from (37) that

$$\left(\frac{\partial E}{\partial T}\right)_V = T \left(\frac{\partial S}{\partial T}\right)_V$$

Since $T dS$ represents heat, the right-hand side gives heat absorbed at constant volume divided by the rise of temperature. This is C_V , the specific heat at constant volume. Consequently

$$\left(\frac{\partial E}{\partial T}\right)_V = C_V \quad (37.2)$$

Similarly

$$\left(\frac{\partial E}{\partial T}\right)_P = T \left(\frac{\partial S}{\partial T}\right)_P - P \left(\frac{\partial V}{\partial T}\right)_P = C_P - P \left(\frac{\partial V}{\partial T}\right)_P \quad (37.3)$$

in which C_P is the specific heat at constant pressure. As for the variation of E with pressure at constant temperature, (37) shows that

$$\left(\frac{\partial E}{\partial P}\right)_T = T \left(\frac{\partial S}{\partial P}\right)_T - P \left(\frac{\partial V}{\partial P}\right)_T$$

or, taking account of (34),

$$\left(\frac{\partial E}{\partial P}\right)_T = -T \left(\frac{\partial V}{\partial T}\right)_P - P \left(\frac{\partial V}{\partial P}\right)_T \quad (37.4)$$

The analogous expression for the change of energy with volume at constant temperature follows similarly from (37), for by that equation

$$\left(\frac{\partial E}{\partial V}\right)_T = T \left(\frac{\partial S}{\partial V}\right)_T - P$$

Combination of this with (35.1) yields

$$\left(\frac{\partial E}{\partial V}\right)_T = T \left(\frac{\partial P}{\partial T}\right)_V - P \quad (37.5)$$

Equation (37.5) has an interesting application to any substance which obeys the gas law:

$$PV = nRT$$

where n is the number of moles of substance. For such a substance

$$\left(\frac{\partial P}{\partial T}\right)_V = \frac{nR}{V} \quad (37.6)$$

When this is substituted in (37.5), the right-hand member becomes equal to $P - P = 0$, which shows that $(\partial E/\partial V)_T = 0$, or in other words that the energy is independent of volume. At the same time we have the relation

$$\left(\frac{\partial E}{\partial V}\right)_T = \left(\frac{\partial E}{\partial P}\right)_T \left(\frac{\partial P}{\partial V}\right)_T$$

It follows, therefore, that the energy is also independent of pressure. Consequently it depends only on temperature. The substance therefore fulfills the two criteria of a perfect gas. It is of interest to observe that the condition that the energy depend only on the temperature is actually less specific than the gas law. It may be easily verified that if we replace the gas law by

$$Pf(V) = RT \quad (37.7)$$

where $f(V)$ is any continuous function of volume, the same result follows. Conversely, if $(\partial E/\partial V)_T = 0$, it can be shown that

$$Pf(V) = RT$$

where R is a constant. For in this case, from (37.5),

$$\frac{T}{P} \left(\frac{\partial P}{\partial T}\right)_V = 1$$

After separation of the variables and integration, this gives

$$\ln T = \ln P + \phi(V)$$

or

$$Pf(V) = RT$$

As a conclusion to the discussion of partial derivatives, we turn finally to the heat content, H . Here there is only one derivative which we need consider. This gives the relation between the temperature coefficient of H and the heat capacity at constant pressure, C_p :

$$\left(\frac{\partial H}{\partial T}\right)_P = C_p \quad (37.8)$$

Equation (37.8) follows at once from the fact that an increment in H , at constant pressure, represents the heat absorbed by the system.

Partial Molal Quantities

The equations we have developed thus far are perfectly general, subject only to the condition that the system to which they are applied be a closed one, that is, that it suffer no interchange of matter with its surroundings during the process under consideration. Although this is a condition that we shall often want to retain, nevertheless, in many applications of thermodynamics, particularly to problems of chemical equilibria and equilibria in multiphase systems, it is desirable to introduce a concept which arises from the consideration of the effect of varying the amounts of the different materials present in a system. This is the concept of partial molal, or partial specific, quantities.

Let us consider a one-phase system, and let X be any extensive property of the system such as volume or energy, i.e., a property whose value is proportional to the total mass of the system. Then X will be a function not only of the several physical variables such as pressure and temperature which define the state of the system but also of the variables n_1, n_2, \dots, n_r , which specify the total amounts of the r components present in it. These latter we shall call the composition variables. Let us now suppose that the physical state of the system is completely defined by the two intensive variables, P and T . Then we can write

$$dX = \frac{\partial X}{\partial P} dP + \frac{\partial X}{\partial T} dT + \sum_{i=1}^r \frac{\partial X}{\partial n_i} \partial n_i \quad (38)$$

The differential coefficients $\partial X / \partial n_i$, P and T being held constant, are called either partial molal or partial specific quantities, depending on whether the n 's are expressed in terms of moles or grams. We shall speak of them hereafter as partial molal quantities and denote them by \bar{X}_i . It should be emphasized that partial molal quantities are always defined in terms of constant temperature and pressure. If other physical variables besides temperature and pressure are required to define the state of the system, then of course it becomes necessary to specify these also in the definition of the partial molal quantities.

The physical meaning of partial molal quantities is best understood from an example. Let us choose as our system a solution containing only two components, salt and water, which we may denote by subscripts 1 and 2, respectively, and let us identify X with the total volume of the solution. Then the differential coefficient \bar{X}_1 , signifying $(\partial V/\partial n_1)_{P,T,n_2} \equiv \bar{V}_1$, is what is called the partial molal volume of the salt in the solution. It represents the increase of volume of the solution, per mole of salt added, when the actual amount added is so small in relation to the total quantity of solution as to leave its composition essentially unchanged. If we plotted the measured volume of the solution against the amount of salt added to a given amount of water, then the quantity $\bar{V}_1 = (\partial V/\partial n_1)_{P,T,n_2}$ would be represented by the slope of the curve. It would, of course, depend at any point on the composition of the solution and on temperature and pressure. It may be thought of as the volume occupied by a mole of salt in the solution, though this is in a sense a fiction. The differential coefficient $\bar{V}_2 \equiv (\partial V/\partial n_2)_{P,T,n_1}$ would represent the same concept applied to the water.

It is an attribute of X , as an extensive property, that if each of the composition variables, n_i , is multiplied by any factor, then X itself is multiplied by the same factor, pressure, temperature, and any other intensive property necessary to define the physical state of the system remaining constant. In mathematical language the same thing may be stated by saying that X is a first-order homogeneous function of the composition variables, n_i . Now according to a mathematical principle known as Euler's theorem, it follows that for such a function, X , we can always write

$$X = \sum_{i=1}^r \frac{\partial X}{\partial n_i} n_i \equiv \sum_{i=1}^r \bar{X}_i n_i \quad (39)$$

This shows how any extensive property may be expressed in integral form in terms of the corresponding partial molal quantities and the composition variables, n_i .

For most purposes, we are not interested in the absolute size of the system, and it is convenient to substitute for the composition variables, n_i , the corresponding mole fractions, N_i , defined by

$$N_i = \frac{n_i}{\sum_{i=1}^r n_i} \quad (40)$$

It will be noted that, since the N 's must all add up to unity, the number of *independent* mole fractions is one less than the number of components.

In order to introduce mole fractions, we may divide both sides of equation (39) by Σn_i . We then obtain, as an alternative expression, denoting by x the value of the property X per mole of total components:

$$x = \sum_{i=1}^r \bar{X}_i N_i \quad (41)$$

It should be observed that the quantity on the left is now no longer an extensive property but the corresponding quantity referred to one mole of material, regardless of how the mole is made up of the various components. For example, in the case of the salt solution, the corresponding volume would be the volume of that amount of solution which is contained in all one mole, partly salt, partly water.

If we differentiate (39) in the most general manner, we obtain at once

$$dX = \Sigma \bar{X}_i dn_i + \Sigma n_i d\bar{X}_i$$

Combination of this with (38) gives

$$\left(\frac{\partial X}{\partial P}\right)_{T,n} dP + \left(\frac{\partial X}{\partial T}\right)_{P,n} dT - \Sigma n_i d\bar{X}_i = 0 \quad (42)$$

This is a fundamental relation between the variations of partial molal quantities, which are, of course, functions of all the variables (P , T , n_i) necessary to define the system. At constant temperature and pressure, (42) becomes simply

$$\Sigma n_i d\bar{X}_i = 0 \quad (43)$$

or, in terms of mole fractions,

$$\Sigma N_i d\bar{X}_i = 0$$

The significance of (43) is brought out by the example of the salt solution considered above, in which X was identified with volume. If we think of N_1 (mole fraction of salt) as the independent variable, we can write

$$N_1 \frac{d\bar{V}_1}{dN_1} = -N_2 \frac{d\bar{V}_2}{dN_1}$$

This shows that if the partial molal volume of the salt increases with addition of salt the partial molal volume of the water must decrease correspondingly. The behavior of one quantity is determined by that of the other.

We offer some comments on the values found experimentally for partial molal volumes and heat capacities. The partial molal volumes of organic compounds in dilute aqueous solutions can generally be expressed

as sums of contributions by the constituent atoms or groups—carbon 9.9 cc/mole, hydrogen 3.1, nitrogen 1.5, hydroxyl oxygen 2.3, sulfur 15.5, and so on. This was shown by J. Traube in the late nineteenth century; he also showed that an additional term for the whole molecule, called the covolume and amounting to approximately 13 cc/mole, had to be added to the sum of the terms for the constituent atoms in order to obtain agreement between calculated and observed partial molal volumes. The covolume is taken as having the same value for all molecules, regardless of their size.

Partial molal volumes of ionic solutes are almost always low, in that the volume occupied by a dilute aqueous solution of a salt is less than the sum of the volumes occupied by the salt and water separately. In some cases, indeed, the partial molal volume of the solute is actually negative. For instance, if a little anhydrous magnesium sulfate is added to pure water, the liquid actually shrinks. The occurrence of such an effect is readily understandable in terms of our picture of the structure of water as presented in Chapter 2. The magnesium ion, with its high charge and small radius, exerts a powerful electrostatic effect in orienting and compressing the surrounding polar water molecules, which become packed closely around it in the intense electrical field. The larger sulfate ion exerts a similar but less pronounced effect. The result is that the volume decrease due to the closer packing of water molecules more than counterbalances the volume increase due to the addition of the magnesium and sulfate ions. This volume decrease is known as electrostriction. It is always found when ionic groups pass into aqueous solution, being most pronounced for ions of high charge and small radius.

Electrostriction effects are also readily recognized in solutions of organic dipolar ions. For instance, the partial molal volume of glycine ($^+\text{H}_3\text{N}\cdot\text{CH}_2\cdot\text{COO}^-$) in water at high dilution is 43.5 cc/mole. On the other hand, the partial molal volume of the isomeric substance glycolamide ($\text{HO}\cdot\text{CH}_2\cdot\text{CONH}_2$) is 56.3 cc/mole. The difference is due to the electrostriction of the water molecules around the two charged groups in the glycine. Similarly, the partial molal volume of α -alanine ($^+\text{H}_3\text{N}\cdot\text{CH}\cdot\text{CH}_3\cdot\text{COO}^-$) is 60.6 cc/mole, while that of its uncharged isomer lactamide ($\text{CH}_3\cdot\text{CHOH}\cdot\text{CONH}_2$) is 73.2 cc/mole. The partial molal volume of β -alanine ($^+\text{H}_3\text{N}\cdot\text{CH}_2\text{CH}_2\cdot\text{COO}^-$) is about 2 cc smaller than that of α -alanine; that is, the electrostriction becomes greater as the oppositely charged groups in the dipolar ion are moved further apart.

Owing to the electrostriction effects mentioned above, the partial molal heat capacities of most salts also have negative values in water; for the electrostriction causes the neighboring water molecules to become more or less rigidly attached to an ion or an ionic group, so that their

motion of translation and rotation is greatly inhibited, with a consequent decrease in the heat capacity of the system. Similar effects are also found on comparing the partial molal heat capacities of dipolar ions with those of their uncharged isomers. Further discussions of these phenomena are given by Cohn and Edsall (1943, Chapters 7 and 16), by Edsall (1953, p. 565 ff) and by Gucker, Klotz and Allen (1942). An excellent general discussion of partial molal properties, and the methods of determining them from experimental data, has been given by Glasstone (1947, Chapter XVIII).

Chemical Potentials

In chemical thermodynamics, much use is made of a variety of partial molal quantities such as partial molal volumes, energies, heats, heat capacities, entropies, and free energies. Of these, perhaps the most important are the partial molal free energies $(\partial F/\partial n_i)_{P,T,n_j}$ which are commonly referred to as chemical potentials and denoted by μ , after Gibbs. It is these which are the decisive quantities in all chemical equilibria. Owing to their great importance, it is worth while to set down together some of the principal relationships involving chemical potentials specifically which result from the general equations just presented.

Since $(\partial F/\partial P)_T = V$ and $(\partial F/\partial T)_P = -S$, equation (38), applied to free energy, is

$$dF = V dP - S dT + \sum \mu_i dn_i \quad (44)$$

where

$$\mu_i = (\partial F/\partial n_i)_{P,T,n_j}$$

The subscripts P , T , n_j indicate that pressure, temperature, and the masses of all components *except component i* are held constant in the process denoted by the differentiation. Similarly (39) and (41) become, respectively,

$$F = \sum \mu_i n_i \quad (45)$$

and

$$F = \sum \mu_i N_i \quad (46)$$

Also (42) becomes

$$-V dP + S dT + \sum n_i d\mu_i = 0 \quad (47)$$

which, at constant temperature and pressure, reduces to

$$\sum n_i d\mu_i = 0 \quad (48)$$

or

$$\sum N_i d\mu_i = 0$$

This last equation, which corresponds to (43) and is called the Gibbs-Duhem relation, is of great significance. In a two-component system,

for example, it enables us to calculate the chemical potential of component 2, if that of component 1 is known over a range of composition.

At this point, it might be interjected that the concept of chemical potential may be given an alternative and equivalent definition to that just offered, as a partial derivative of energy instead of free energy. Actually this is the definition originally proposed by Gibbs. Gibbs' definition may be obtained at once from ours by substituting into (44) the definition

$$F = E + PV - TS$$

in differential form. This yields the equation

$$dE = T dS - P dV + \sum \mu_i dn_i \quad (49)$$

which expresses E as a function of S , V , and the n_i . It follows that

$$\mu_i \equiv \left(\frac{\partial F}{\partial n_i} \right)_{P,T,n_j} = \left(\frac{\partial E}{\partial n_i} \right)_{S,V,n_j} \quad (50)$$

Still other definitions of μ_i are possible, in terms of the heat content, H , and the quantity $A = E - TS$, sometimes called the "Helmholtz free energy":

$$\mu_i = \left(\frac{\partial H}{\partial n_i} \right)_{S,P,n_j} = \left(\frac{\partial A}{\partial n_i} \right)_{T,V,n_j}$$

We mention these definitions of μ_i , however, merely to show that the same chemical potential may be defined in terms of several sets of independent variables. In chemical, and especially in biochemical, practice, it is generally most convenient to take pressure and temperature as the independent variables, and we define μ_i experimentally as $\partial F / \partial n_i$ by (44).

Before we pass on to applications of the concept of chemical potential, it is worth while to obtain expressions for its variation with pressure and temperature. These follow at once from (44). Since dF expressed as a function of dP , dT , and the dn_i is an exact differential, the second-order cross derivatives, e.g., $\partial^2 F / \partial P \partial n_i$ and $\partial^2 F / \partial n_i \partial P$, are equal to one another. Consequently

$$\left(\frac{\partial \mu_i}{\partial P} \right)_{n,T} = \left(\frac{\partial V}{\partial n_i} \right)_{P,T} = \bar{V}_i \quad (51)$$

Similarly

$$\left(\frac{\partial \mu_i}{\partial T} \right)_{P,n} = -\bar{S}_i \quad (52)$$

Thus the familiar relations applicable to free energy itself carry over to chemical potentials except for the distinction between total and partial molal quantities.

Phase Rule

The concept of chemical potential finds immediate application in the study of the equilibrium between two or more phases. The condition of equilibrium is that for any infinitesimal change of the system consistent with the prevailing constraints the change of free energy be zero, the temperature throughout the system and the pressure in each phase remaining constant. Now the only constraint imposed on the system is that it be closed, that is, that there be no net gain or loss in the total amount of any component in the system as a whole, although there may, of course, be exchanges between different phases. Clearly this demands that the chemical potential of each component be the same in all phases. If it were different for any component in two phases, 1 and 2, then as a result of the passage of dn moles of that component from one of these phases to the other there would be a change of free energy,

$$dF = dn(\mu_2 - \mu_1)$$

different from zero. This is contrary to the condition of equilibrium.

This result provides the basis of the phase rule, which we shall now formulate as it applies to a system of any number of phases all at the same pressure as well as the same temperature. Since the chemical potentials in each phase are functions of the variables which define the state of the phase, the equality of the potentials imposes restrictions on these variables. Suppose that the number of components in the system is β , and that the number of phases is α . Then the number of composition variables required to characterize each phase will be $\beta - 1$, since, of course, it is only the relative amounts of the different components that matter. The total number of composition variables applicable to the whole system (α phases) will, therefore, be $\alpha(\beta - 1)$, and the total number of variables, including temperature and pressure, will be $\alpha(\beta - 1) + 2$. At the same time, for each component, the equality of the potentials in the α phases involves $\alpha - 1$ independent relations between the variables, one phase serving as a reference phase. The total number of restrictions arising from all the β components is, therefore, $\beta(\alpha - 1)$. The difference between the number of variables and the number of restrictions gives the number of independent variables. This is referred to as the number of degrees of freedom of the system. It follows from what has just been said that the number of degrees of freedom, taking account of pressure and temperature, is

$$\alpha(\beta - 1) + 2 - \beta(\alpha - 1) = 2 + \beta - \alpha \quad (53)$$

This relationship between the number of phases, number of components, and number of degrees of freedom is called the phase rule. It was first recognized by Gibbs.

It is worth while to discuss one or two applications of this fundamental principle. Let us first consider a system containing only one component but two phases, such as liquid water in contact with ice. Then the phase rule tells us immediately that there is only one degree of freedom. Since there is only one component, no composition variables are involved. It follows, therefore, that the pressure is uniquely determined by the temperature, and vice versa. We have already discovered the nature of the dependence in the Clausius-Clapeyron relation embodied in equations (31) and (32).

In the case where a third phase, e.g., the vapor phase, is added, the number of degrees of freedom is reduced to zero. That is to say, the pressure and temperature are uniquely fixed. They determine what is called the triple point. In the case of water, the triple point occurs at a pressure of 4.5 mm Hg and a temperature of $+0.0075^{\circ}$.

The triple point may be represented as a point on a graph of which the ordinates and abscissae give pressures and temperatures, respectively. The equilibrium of two phases will then be represented by a line on the graph passing through this point whose slope is everywhere determined by the Clausius-Clapeyron relationship. There will be three such lines, one for each of the three ways of choosing two phases from three, and these lines will all meet at the triple point. Such a graph, representing the equilibria between the various phases, is called a phase diagram. It often happens that a substance in the solid state is capable of existing in several distinct allotropic forms. For example: solid water exists in six such allotropic forms of which ordinary ice is one. In such cases, the phase diagram becomes quite complex, as shown by Fig. 1, which represents water. It still remains true, however, that only three lines intersect at a point, since not more than three phases can exist together. This follows from the fact that the coexistence of three phases reduces the degrees of freedom of a one-component system to zero. There are, therefore, a number of triple points for a substance like water.

There are several general features of the phase diagram of a one-component system which are worth pointing out before we pass on to multicomponent systems. If we consider any line on such a diagram, then the phase which is stable on the high-temperature side of the line is the phase of higher heat content. This follows directly from the fact that at a point on the line the free energies of the two phases are necessarily equal, since the line represents an equilibrium, but that on either side of the line the stable form must have a lower free energy than the other. Since the rate of change of free energy with temperature (at constant pressure) is $-S$, it follows that in order that the high-temperature form may have a lower free energy than the other at any point to the right

of the line it must have a higher entropy and consequently a greater heat content. Similarly the form stable on the high-pressure side of such a line is the form of smaller volume. This follows by a strictly analogous piece of reasoning, taking account of the fact that $(\partial F/\partial P)_T = V$.

It is an experimental fact that, of two phases in equilibrium, the phase of higher heat content is generally also the phase of larger volume, and whenever this is so it follows from equation (32) that the corresponding line on the phase diagram must have a positive slope. There are, however,

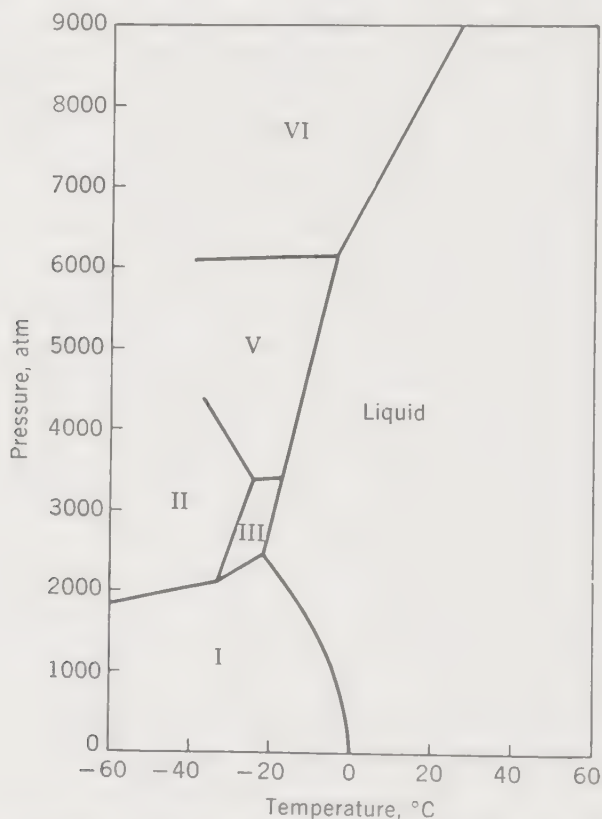


FIG. 1. Phase diagram of water and of various types of ice, as a function of pressure and temperature.

rare cases, of which liquid water in equilibrium with ordinary ice is one, in which the opposite is true. In such cases the corresponding line on the phase diagram has a negative slope. Whatever the slope of the lines on a phase diagram, however, it is always true that any line produced through the triple point lies between the other two lines, as illustrated in Fig. 2, where the three phases are denoted by subscripts 1, 2, and 3. Since, by (32),

$$\frac{dP_{13}}{dT} = \frac{H_3 - H_1}{T(V_{3,1} - V_1)}$$

and similarly for the other lines, this means that if

$$\frac{H_2 - H_1}{V_2 - V_1} > \frac{H_3 - H_1}{V_3 - V_1} \quad (54)$$

then

$$\frac{H_3 - H_1}{V_3 - V_1} > \frac{H_3 - H_2}{V_3 - V_2} \quad (55)$$

and conversely. A proof of this is as follows. On the basis of the principles of the last paragraph, it will be seen for the case illustrated above that

$$V_3 > V_2 > V_1$$

It follows from this that both the products $(V_2 - V_1)(V_3 - V_1)$ and $(V_3 - V_2)(V_3 - V_1)$ are positive quantities. We may, therefore, clear the two expressions (54) and (55) of fractions without affecting the in-

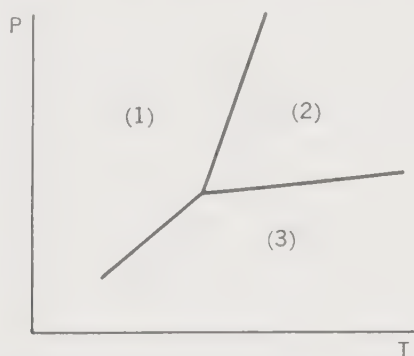


FIG. 2. One of the possible types of intersection of the pressure-temperature curves in a one-component system near the triple point.

equalities. When this is done, however, they become identical, as may be easily verified, which shows that either expression may be derived from the other.

In the case of a multicomponent system the number of variables is increased by the composition variables, and it becomes possible for more than three phases to exist together, e.g., solid salt, ice, an aqueous salt solution, and vapor. The question may immediately be raised as to whether in such cases the basic argument underlying the phase rule is applicable, since the solid phases often contain only one of the components, as does ice. If so, then no composition variable is required to fix the state of the phase (e.g., ice), as was presupposed in the argument. Reflection shows, however, that the difficulty here is illusory: if one component is excluded from one of the phases, it is true that the number of variables is diminished by one, but so too is the number of restrictions, since the chemical potential of the excluded component in the phase in

question does not come into play. In determining the number of degrees of freedom of a system, we may, therefore, always proceed as if each component were present in each phase and apply the phase rule as it stands.

There is a question that may have been a source of perplexity to the reader in regard to the triple point of water which may appropriately be discussed here. It was pointed out that this point corresponds to a temperature slightly above zero (0.0075°), where the vapor pressure of water is 4.5 mm. Yet it is well known that under ordinary conditions the melting point of ice is exactly zero degree, this temperature being in fact defined in terms of the equilibrium of ice, liquid water, and water vapor. The answer to the apparent inconsistency is that by "ordinary conditions" we mean that the ice-water system is exposed to the atmosphere. This means that the system is no longer a one-component system, and that we have the additional degrees of freedom and additional restrictions associated with the presence of all the additional components contributed by the air, e.g., oxygen, nitrogen, argon, carbon dioxide, and traces of other vapors. The presence of these components is bound to affect the equilibrium. Let us denote their number by r . Then the total number of components, taking account of water, is $r + 1$, and the phase rule tells us that the number of degrees of freedom is $(r + 1) + 2 - 3 = r$. The composition of dry air, however, is fixed and determines the values of $r - 1$ composition variables in the vapor phase. Hence, the actual number of degrees of freedom is really only $r - (r - 1) = 1$. But if the total pressure is also fixed, at one atmosphere, even this one variable is exhausted and the temperature is uniquely determined. Under these conditions, it is zero, different, of course, though only slightly different, from what it is in the absence of the air. Physically the difference may be attributed in part to the effect of hydrostatic pressure on the two condensed phases (ice and water) and in part to the effect of gases dissolved in the liquid phase.

The number of components in a system, it should be emphasized, is the number of chemical entities whose amounts in the system may be fixed independently of one another. Sometimes there is a question as to whether or not there may be a chemical equilibrium, representing a reaction, between certain substances. If so, of course, the number of components in a system containing the substances is reduced by one, since, without going into detail, it is clear that the condition of equilibrium must determine the relative amounts of the substances. Whether or not such equilibrium prevails may, therefore, be determined on the basis of the phase rule by exploring the number of degrees of freedom of the system. This is illustrated by the case of a solid amino acid, say

glycine, in contact with gaseous hydrochloric acid in a vapor phase. The question is, does the hydrochloric acid react with the amino acid to produce solid glycine hydrochloride as a new phase? If so, the phase rule tells us that there will be only one degree of freedom. If, therefore, we fixed the temperature of a gas-tight chamber containing the amino acid and a gas phase of hydrochloric acid, then the pressure would remain constant regardless of the amount of the hydrochloric acid introduced until at last the pure glycine phase was exhausted. On the other hand, if there were no reaction, the pressure would increase on addition of gaseous hydrochloric acid according to the gas laws.⁷ Experiment shows that it is the former alternative which is in fact realized. The use of the phase rule to settle such questions as this is of considerable importance.

Activities

It is easy to see that if a volatile component is present in two phases equilibrium demands that its vapor pressure be the same in each. Otherwise, if the two phases were brought into contact with a vapor phase the component would distill over from one phase to the other. But equilibrium also, of course, demands an equality of chemical potentials. Vapor pressure must, therefore, be a very direct expression of chemical potential. The relation between the two is a matter we shall now consider. It leads us directly into the concept of activities.

Let us begin with the simplest case, that of a pure gas. Since, as we have shown, $(\partial F/\partial P)_T = V$, it follows that at any one temperature the free energy of the gas is given as a function of pressure by

$$F = \int V dP$$

If we assume the gas to be ideal, then this becomes

$$F = \int \frac{nRT}{P} dP = nRT \ln P + \text{const.} \quad (56)$$

in which n denotes the number of moles under consideration. The constant of integration is determined by any pair of values of F and P . We have denoted the total pressure on a system or phase by the symbol P , using p to denote the vapor pressure of a component of the system. For a pure one-component system, $p = P$, if the vapor phase is present alone, or in equilibrium with the liquid or solid phase or both. We therefore denote the pressure in a one-component gas phase by p , to emphasize the analogy with the equations for systems of two or more

⁷ The same thing would be true if there were a reaction but the glycine hydrochloride formed a solid solution in the glycine.

components, which will be discussed below. On this basis we rewrite (56) as:

$$F - F^\circ = nRT \ln \frac{p}{p^\circ}$$

where F° is the value of F at $p = p^\circ$. For a pure substance the chemical potential $(\partial F/\partial n)_{P,T}$ is the same as the free energy per mole. The above relation tells us, therefore, that the chemical potential of an ideal gas may be expressed as

$$\mu - \mu^\circ = RT \ln \frac{p}{p^\circ} \equiv RT \ln a \quad (57)$$

The quantity a defined by (57) is called the activity of the gas. By definition the state of unit activity is the standard state. Clearly this corresponds to $p = p^\circ$ and $\mu = \mu^\circ$; p° and μ° are, therefore, the pressure and chemical potential respectively of the gas in the standard state. We are, of course, at liberty to choose the standard state as we please. It should be emphasized that the chemical potential, like the free energy from which it is derived, contains an undetermined constant, although in the difference $\mu - \mu^\circ$ this constant cancels out.

Equations (56) and (57) were derived for an ideal gas. For a gas which is not ideal, the integral $RT \int V dp$ is no longer exactly equal to $RT \ln (p/p^\circ)$. The integral may, however, always be formulated as $RT \ln p$ plus a function of p . That this is so is evident if we express V as

$$V = \frac{RT}{p} + \phi(p)$$

which is surely possible. It follows from this that for an imperfect gas equation (57) is to be replaced by the equation

$$\mu - \mu^\circ = RT \ln \frac{f(p)p}{p^\circ} \equiv RT \ln a \quad (58)$$

In so far as this equation relates activity to chemical potential it is the same as (57), and indeed constitutes the universal definition of activity; on the other hand, as regards the relation of activity and chemical potential to pressure, it is different owing to the presence of the factor $f(p)$. It is generally only in gases at high pressure that $f(p)$ differs much from unity. In the systems with which the biochemist deals, however, the pressures commonly encountered are of the order of one atmosphere, or less, and it is generally unnecessary to take account of departures from the ideal gas law in discussing the chemical potentials of vapors, and we may take activity as proportional to vapor pressure without significant error. We shall do this hereafter, applying equation (57).

The question next arises as to the chemical potential of a gas in a mixture of gases. We shall show that this is related to the partial pressure

of the gas in the same way that the chemical potential of a pure gas is related to the total pressure. The partial pressure of a gas in a mixture is equal to the mole fraction of the gas in the mixture times the total pressure. This is what is called Dalton's law. In a sense, however, it is no law at all but only a definition, and the partial pressure is a fiction, since we have no way of measuring separately the pressure due to one component only. It is true, to be sure, that the total pressure calculated by assigning to each gas the pressure it would exert in accordance with the gas law in the absence of all other gases, and by taking account of the amounts of all the gases present, agrees with the observed pressure. But this would hardly justify the extension of equation (57) or (58) to partial pressures were it not for a further experimental fact of great significance. This involves the behavior of hydrogen in a mixture of gases separated from pure hydrogen by a piece of hot platinum foil. Under these conditions the hydrogen passes readily through the foil while the other gases are held back. The foil acts as a semipermeable membrane for hydrogen. Now at equilibrium it is found that the pressure of the hydrogen in the phase where it exists alone is the same as its calculated partial pressure in the mixture of gases constituting the other phase. Since at equilibrium the chemical potential of the hydrogen must be the same in both phases, and since we know that equation (57) is valid for the pure phase, it follows that the chemical potential of the hydrogen in the mixture is related to its partial pressure according to equation (57) or (58). This holds for hydrogen. Unfortunately no other membranes selectively permeable to other gases are known. Nevertheless there is nothing to prevent our extending the principle to all other gases, and in fact there is every practical reason to do so. We may, therefore, treat partial pressures like total pressures in reckoning activities and chemical potentials.

We turn now to the question of the activities and chemical potentials of components present in solid and liquid phases, that is to say, in condensed phases. Here again vapor pressure is of fundamental importance. The partial pressure of a volatile component in a gas phase in equilibrium with a condensed phase is spoken of as the vapor pressure of the component *in* the condensed phase and is taken as equal to or proportional to its activity there. In the light of what has been said, this is a perfectly justifiable procedure, owing to the necessary equality of chemical potentials in all phases at equilibrium, provided only that we are content to treat the vapor as a perfect gas. We may, therefore, apply equation (57) to calculate the chemical potential of a volatile component in a condensed phase, although vapor pressure is not a property which can be determined from measurements on the condensed phase itself. Even when we are concerned with volatile components, it is the aim of a

thermodynamic treatment to relate chemical potentials, as determined from vapor pressures in the gaseous phase, to the composition of the liquid phase, often most rationally expressed in terms of the mole fractions of the components. In dealing with nonvolatile components, such as salts, amino acids, or proteins, we must find other ways of determining the chemical potentials. First, however, we consider systems of volatile components.

Ideal (or Perfect) Solutions and Raoult's Law

The most convenient point of reference in our discussion is what G. N. Lewis has called the ideal solution, in which the vapor pressure and therefore the activity of each component, at a given pressure and temperature, is *proportional* to its mole fraction in the solution. If the pure component is chosen as the standard state, then its activity is equal to its mole fraction. Let the vapor pressure of pure component i , at the given pressure and temperature, be p_i° . Then in an ideal solution, if the mole fraction of component i is N_i , its vapor pressure p_i is $N_i p_i^\circ$ and its activity is $(p_i/p_i^\circ) = N_i$. No actual solution can be expected to be rigorously ideal, just as no actual gas is a perfect gas, but some solutions of very closely related substances approach the ideal—for instance, benzene and toluene, or ethylene bromide and propylene bromide. The form of the curve for the partial vapor pressures in a two-component ideal solution is shown in Fig. 3. The curves for the two components, as a function of the mole fraction of either component, are straight lines, and the total vapor pressure is also a linear function of composition, as shown by the dotted line.

A further characteristic of an ideal solution is that the components must mix without change of volume, and without absorption or evolution of heat; that is, $\Delta V = 0$ and $\Delta H = 0$ for the mixing process. This is to be expected, since in an ideal solution the different kinds of molecules must be so similar that any one molecule in the solution interacts with one of another species in essentially the same way as it interacts with one of its own kind. If the solution is to remain ideal over a range of pressure and temperature, it can be shown from thermodynamic principles—see equations (62) and (63) below—that ΔV and ΔH of mixing must be zero over this range.

The molal chemical potentials of the components of an ideal solution are given by the expressions

$$\mu_1 - \mu_1^\circ = RT \ln N_1 \quad (59)$$

$$\mu_2 - \mu_2^\circ = RT \ln N_2 \quad (60)$$

If we mix N_1 moles of component 1 with N_2 moles of component 2, to make one mole of mixture, then the free energy of mixing, per mole of mixture, is

$$\Delta F_{\text{ideal}} = N_1(\mu_1 - \mu_1^\circ) + N_2(\mu_2 - \mu_2^\circ) = N_1RT \ln N_1 + N_2RT \ln N_2 \quad (61)$$

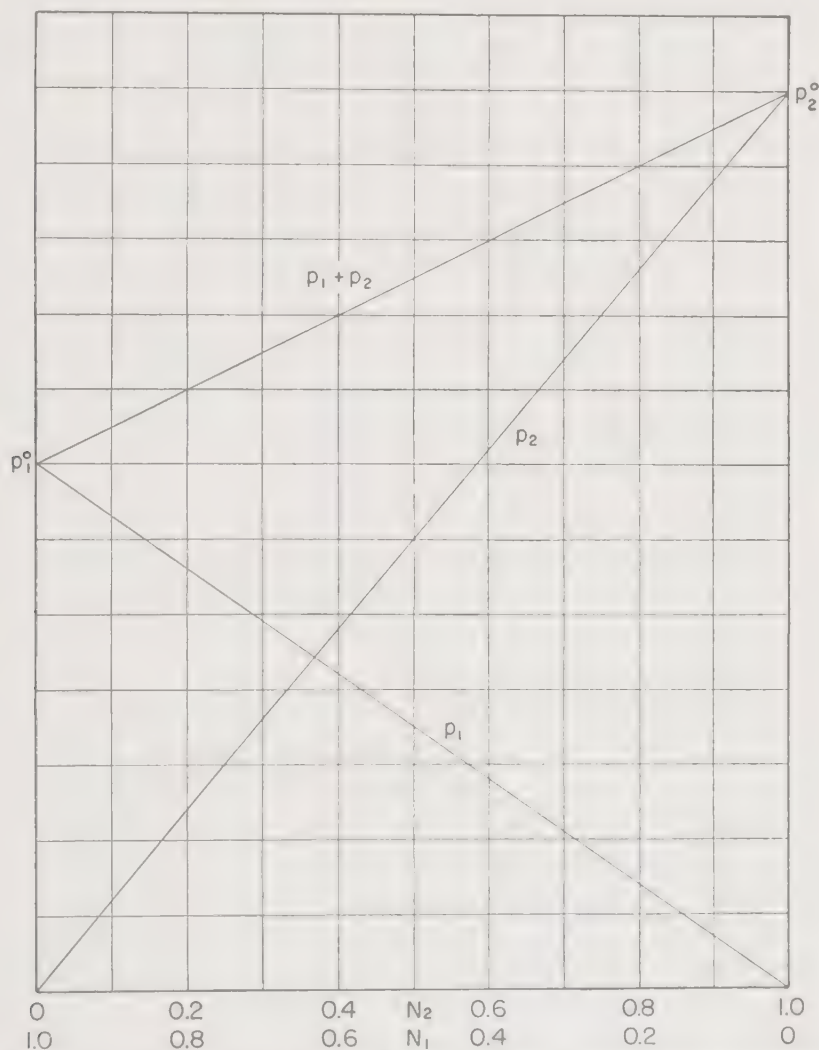


FIG. 3. The vapor pressure diagram for an ideal solution of two components. The vapor pressure of each component is proportional to its mole fraction in the solution.

Since N_1 and N_2 are necessarily less than unity, ΔF is always negative. This must, of course, be true, since the mixing is a spontaneous process at constant P and T .

Equation (61) holds for all values of temperature and pressure for which the solution is ideal. Consequently $\Delta F_{\text{ideal}}/T$ is independent of

pressure and temperature. Since, however, ΔF for any process must obey the relations

$$\left(\frac{\partial \Delta F}{\partial P}\right)_T = \Delta V \quad (\text{corresponding to 30.1}) \quad (62)$$

and

$$\left(\frac{\partial \frac{\Delta F}{T}}{\partial T}\right)_P = -\frac{\Delta H}{T^2} \quad (\text{corresponding to 30.2}) \quad (63)$$

it follows that there can be no heat liberated and no volume change on the formation of the ideal solution. Therefore, since $\Delta H = 0$ for the process of mixing the components of an ideal solution, ΔF must be equal to $-T \Delta S$, and the entropy of mixing is given by:

$$\Delta S_{\text{ideal}} = -\Delta F/T = -N_1 R \ln N_1 - N_2 R \ln N_2 \quad (64)$$

which is always positive, and is independent of temperature.

The equation for the vapor pressure or activity of a component in an ideal solution is

$$a_i = p_i/p_i^\circ = N_i \quad (65)$$

This represents Raoult's law when the component is regarded as the solvent and Henry's law when it is regarded as the solute. It is obeyed approximately by a few actual systems over the whole range of composition, as we have mentioned above. It is of great importance however that Raoult's law is valid, *as a limiting law*, quite generally for a component i provided N_i is close to unity; that is, for very dilute solutions in which component i is the solvent. This principle, which is by no means self-evident, follows by the Gibbs-Duhem relation from the obvious fact that the vapor pressure and consequently the activity of any component, at sufficiently high dilution, is proportional to its mole fraction—a restricted form of Henry's law. This is justifiable, because any continuous function can be treated as linear over a sufficiently small range. In this case the function, p , is zero when $N = 0$. Consequently p is proportional to N , in a small range starting with $N = 0$. Consider a two-component system. By (57), for either component,

$$d(\mu_i - \mu_i^\circ) = d\mu_i = RT d \ln a_i \quad (66)$$

Therefore, by (48):

$$n_1 d \ln a_1 + n_2 d \ln a_2 = 0 \quad (67)$$

If we set $a_i = p_i/p_i^\circ$ and divide by $n_1 + n_2$ this becomes

$$N_1 d \ln p_1 + N_2 d \ln p_2 = 0 \quad (68)$$

But, since $dN_1 = -dN_2$, this gives:

$$N_1 d \ln p_1 / dN_1 = N_2 d \ln p_2 / dN_2 \quad (69)$$

or

$$d \ln p_1 / d \ln N_1 = d \ln p_2 / d \ln N_2$$

If we identify component (2) with the solute, present at sufficiently high dilution so that p_2 is proportional to N_2 , then the right hand member of (69) must be equal to unity. Hence also:

$$N_1 (d \ln p_1 / dN_1) = 1$$

Integration of this gives

$$p_1 = p_1^\circ(N_1) \quad (N_1 \rightarrow 1, N_2 \rightarrow 0)$$

p_1° , which represents the constant of integration, is of course the vapor pressure of the pure component (1). This is Raoult's law for component (1), i.e. for the solvent, in a dilute solution.

We may now consider a solution which deviates markedly from ideal behavior. As an example we choose the system water (component 1)-*n*-propanol (component 2). The partial vapor pressures for this system at 25° are represented in Fig. 4. The dotted lines indicate the partial vapor pressure curves that would be expected for an ideal solution; the actual data are shown by the solid lines, and it is seen that they lie everywhere above the ideal curves, except that the curve for p_1 becomes tangent to the ideal curve for component 1 as $N_1 \rightarrow 1$, and the curve for p_2 becomes tangent to the ideal curve for component 2 as $N_2 \rightarrow 1$. This geometrical fact is an expression of Raoult's limiting law (equations 59 and 60). Over most of the domain covered by the diagrams of Fig. 4, however, the experimental curves lie well above the ideal curves; that is, the deviations from Raoult's law are positive. In terms of molecular interactions this means that water molecules tend to associate with other water molecules, and propanol molecules with other propanol molecules; thus each kind of molecule tends to attract its own kind and, at least relatively, to repel the other kind. Hence, when the two are mixed, each increases the activity of the other above what it would be in an ideal solution at the same mole fraction. Water-methanol and water-ethanol systems also show positive deviations from Raoult's law; they are least marked for methanol and increase progressively with the length of the hydrocarbon chain of the alcohol. In the water-*n*-butanol system the positive deviations from Raoult's law become so great that the mixture separates into two phases, over a wide range of composition—an upper phase rich in butanol and a lower phase rich in water.

Other systems are known which show negative deviations from Raoult's law; the actual vapor pressure curves lie *below* the straight-line relations for an ideal solution. An example is the chloroform-acetone system. The negative deviation from Raoult's law in this case may be explained by the evidence from other sources that the one hydrogen atom in chloroform has a strong tendency to form hydrogen bonds, due

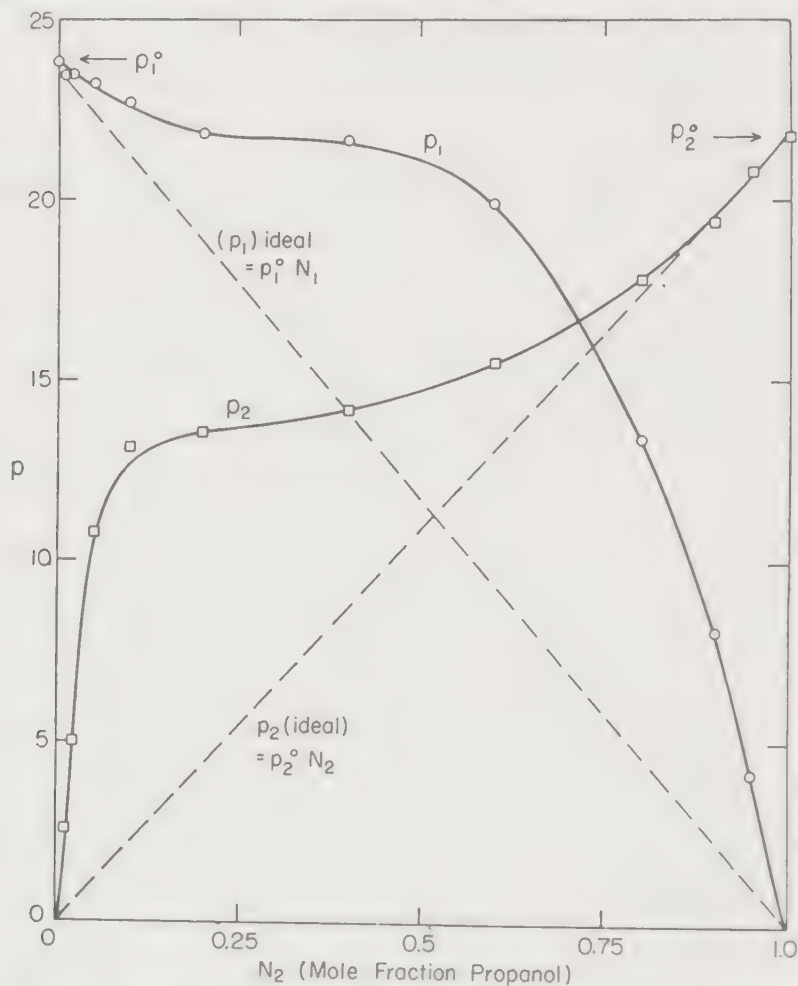


FIG. 4. Vapor pressure diagram for the system water-*n*-propanol at 25°. (From data of J. A. V. Butler, D. W. Thomson, and W. H. Maclellan (1933). *J. Chem. Soc.* p. 674.) The dotted lines show the partial vapor pressures that would be obtained if the solution were ideal.

to the very polar C^+-Cl^- bonds, which draw electrons away from the carbon atom. Thus the hydrogen atom in chloroform is attracted strongly to the oxygen end of the highly polar $C=O$ group in the acetone molecule, forming a readily dissociable complex, which may be written $Cl_3CH \cdots O=C(CH_3)_2$. The formation of this complex lowers the activity, and hence the vapor pressure, of the free acetone and chloroform

molecules in the solution. For a valuable and extensive further discussion of such phenomena, the reader may consult Hildebrand and Scott (1950).

These considerations have carried us beyond the realm of pure thermodynamics onto a more molecular level, a step which we shall often take in endeavoring to discover what thermodynamic data mean in terms of molecular structure and interactions. From the point of view of pure thermodynamics, however, we need no explanations in terms of molecular mechanisms; as long as the experimental data are reliable we may insert them into calculations of entropy or free energy changes without asking *why* they happen to have the particular values that are found experimentally.

We now return to purely thermodynamic considerations. Equation (69) shows how in any two-component system, such as the water-propanol system, the two vapor pressure curves are interrelated over their whole course. The slope of the curve for p_1 as a function of N_1 must be opposite in sign to that for p_2 as a function of N_1 , and the ratio of the two slopes must be $-N_2p_1/N_1p_2$. This serves as an important check on the experimental data; if the experimental slopes, for any region of composition on the curves, are not related by (69), we may infer an experimental error in the vapor pressure determinations. Moreover it is apparent that, if we knew the vapor pressure of one component, say p_1 , as a function of composition, and knew a single point, say the value of p_2° at $N_2 = 1$, for the other component, it would be possible to trace out the entire curve for the second component, since equation (69) defines the slope of the curve at every point. This is in principle exactly what is done to obtain the activity of component 2 from that of component 1, if component 2 has a vapor pressure too low to be measured, as for a salt or an amino acid. We shall return shortly to the details of procedure in such cases.

Activity Coefficients and the Choice of Standard States

For non-ideal solutions it is convenient to describe the deviations of the components from ideal behavior in terms of what is generally called an activity coefficient. This is defined as the ratio of the activity to the mole fraction (or sometimes to concentration expressed in some other way). Thus, if we denote the activity coefficient by f ,

$$a_i = f_i N_i \quad (70)$$

The activity coefficient is dependent on the choice of the standard state. The choice of the standard state, in which the activity is defined as unity, may be made to suit our convenience. One principal convention, which we have employed in the preceding discussion, is to take the pure compo-

ment, at the same temperature and total pressure as that of the experiment, as the standard state. By this convention, therefore:

$$f_i = a_i/N_i = p_i/p_i^\circ N_i \quad (70.1)$$

The value of f_i is equal to one in any solution where N_i is so close to unity that Raoult's law holds, for by that law, $p_i = p_i^\circ N_i$. Of course in any ideal solution, where Raoult's law holds for all concentrations, the activity coefficient, in accordance with this convention, is equal to unity for all values of N (i.e. at all concentrations).

Another frequency employed convention for defining the standard state of a component consists in taking its activity as *equal* to its mole fraction in an infinitely dilute solution where its mole fraction approaches zero and where Henry's law, as a limiting law, is applicable. This is justified, since according to this law the vapor pressure is proportional to the mole fraction. This convention has, of course, the effect of making the activity coefficient equal to unity in the range where Henry's law is valid. It is important to consider what this convention implies as to the standard state itself. Let us write Henry's law for component i as:

$$p_i = k_i N_i$$

Then, in the range where the law is valid,

$$a_i' = p_i/p_i^* = k_i N_i/p_i^* = N_i \quad (\text{as } N_i \rightarrow 0)$$

It follows that $k_i = p_i^*$. This means that p_i^* , the vapor pressure of the component in the standard state, is the same as the ideal vapor pressure which the component *would have* in the pure state ($N_i = 1$) if Henry's law, valid in the region of infinite dilution, were applicable over the whole range of concentration. For any but an ideal solution this will of course be different from the actual vapor pressure of the pure component, and it may or may not correspond to any realizable state, i.e. to the vapor pressure of any actual solution containing the component in question. For instance, for the water-propanol system shown in Fig. 4, we can determine p_2^* by drawing a tangent to the limiting portion of the curve for p_2 as a function of N_2 , in the range where N_2 approaches zero, and extrapolating this tangent line until it cuts the ordinate axis for $N_2 = 1$. The point of intersection on this ordinate gives p_2^* , the "standard vapor pressure" according to this convention. In this case p_2^* , as we shall see shortly, is much greater than p_2° , and therefore corresponds to no state which is actually attainable. An exactly analogous procedure can be employed to determine p_1^* for water in the propanol-water system: this is also found to be considerably greater than p_1° . Indeed in any system for

which the components show positive deviations from Raoult's law, we may expect to find that $p_i^* > p_i^\circ$, for any component i .

The activity coefficient, for component i , according to this second convention, is (see also equations (73) and (74) below):

$$f_i' = p_i/k_i N_i \equiv p_i/p_i^* N_i \quad (70.2)$$

Thus, at any given concentration of component i , the ratio of the activity coefficients f_i (from 70.1) and f_i' (from 70.2) is:

$$f_i/f_i' = p_i^*/p_i^\circ \equiv k_i/p_i^\circ \quad (70.3)$$

TABLE I
VAPOR PRESSURES AND ACTIVITY COEFFICIENTS IN WATER-PROPANOL MIXTURES

N_2	Water ($p_1^\circ = 23.8$; $k_1 = 3.7p_1^\circ$)			Propanol ($p_2^\circ = 21.8$; $k_2 = 14.4p_2^\circ$)		
	$p_1/p_1^\circ = a_1$	$f_1 = p_1/p_1^\circ N_1$	$f_1' = p_1/k_1 N_1$	$p_2/p_2^\circ = a_2$	$f_2 = p_2/p_2^\circ N_2$	$f_2' = p_2/k_2 N_2$
0.00	1.00	1.000	0.270	—	(14.4)	1.000
0.01	(0.984)	(0.994)	(0.268)	0.123	12.3	0.854
0.02	(0.988)	(1.008)	(0.272)	0.232	11.6	0.805
0.05	0.974	1.025	0.277	0.495	9.9	0.69
0.10	0.954	1.060	0.286	0.605	6.05	0.420
0.20	0.916	1.145	0.309	0.624	3.12	0.217
0.40	0.912	1.52	0.410	0.652	1.63	0.113
0.60	0.836	2.09	0.564	0.712	1.19	0.083
0.80	0.563	2.82	0.762	0.816	1.02	0.071
0.90	0.340	3.40	0.92	(0.89)	(0.986)	(0.0695)
0.95	0.176	3.52	0.95	(0.955)	(1.005)	(0.070)
1.00	—	(3.7)	1.000	1.000	1.000	0.0694

From the data of J. A. V. Butler, D. W. Thompson, and W. H. MacLennan (1933). *J. Chem. Soc.* p. 674.

This ratio is, of course, a constant, independent of N_i although both f_i and f_i' are in general functions of N_i and of the composition of the system as a whole.

The water-propanol system (Fig. 4) is, of course, far from ideal. Table I gives the activity coefficients, at various compositions, for the components of this system corresponding to each of the above conventions. It is apparent that the activity coefficient of each component rises as the mole fraction of the other component increases; by the definition of f given in (70.1) the activity coefficient of water in pure propanol is estimated by extrapolation as 3.7, and the activity coefficient of pro-

panol in pure water as 14.4, although these estimates are not of very high precision. The limiting slope for the vapor pressure curve of propanol in water may thus be written

$$p_2 = k_2 N_2 = 14.4 p_2^\circ N_2 \quad (\text{as } N_2 \rightarrow 0) \quad (71)$$

and the limiting slope for the curve for water in propanol may be written

$$p_1 = k_1 N_1 = 3.7 p_1^\circ N_1 \quad (\text{as } N_1 \rightarrow 0) \quad (72)$$

We may note that the water-propanol system also deviates markedly from ideality with respect to the heat and volume changes that occur on mixing. A mixture of water and propanol occupies a smaller volume than the two components occupied separately before mixing. A solution containing one mole of propanol dissolved in a very large volume of water occupies a volume approximately 12 cc less than the sum of the volumes of the propanol and water before mixing. In the process of forming such a solution, at constant pressure and temperature, a large amount of heat is evolved—approximately $\Delta H = -3350$ cal/mole of propanol. In an ideal solution, of course, ΔH would be zero. The entropy of solution of propanol in water given by

$$T\Delta S = \Delta H - \Delta F$$

thus contains a large negative term, as compared with the entropy change for an ideal solution at the same mole fraction. If we compare the molal entropy of transfer of pure propanol ($N_2 = 1$) to a very dilute solution of propanol in water (N_2 finite but close to zero) with the entropy of transfer for an ideal solute at the same final mole fraction, we obtain

$$T(\Delta S - \Delta S_{\text{ideal}}) = \Delta H - (\Delta F - \Delta F_{\text{ideal}}) = \Delta H - RT \ln f_2$$

In a very dilute solution ($N_2 \rightarrow 0$), f_2 is 14.4 from Table I, and $RT \ln f_2 = 1580$ cal/mole. Since $\Delta H = -3350$, this gives

$$\Delta S - \Delta S_{\text{ideal}} = -16.6 \text{ cal deg}^{-1} \text{ mole}^{-1}$$

Such deviations from ideality are typical of aqueous solutions of most organic compounds, the deviations generally becoming large if the organic compound contains large hydrocarbon residues, and being relatively small if it contains many hydroxyl groups, like the sugars.

We now consider values of f_1' and f_2' calculated by the second convention for choosing the standard state. These are given by

$$f_1' = p_1/k_1 N_1 = f_1 p_1^\circ/k_1 \quad (73)$$

$$f_2' = p_2/k_2 N_2 = f_2 p_2^\circ/k_2 \quad (74)$$

The standard state of unit activity by this second convention, it will be recalled, is the state in which the vapor pressure of component i is equal

to k_i , the proportionality constant in Henry's law, for the given component in very dilute solution (equation 70.2).

Thus the standard state for propanol, by this convention, is a hypothetical state in which the vapor pressure is 14.4 times as great as the vapor pressure of pure propanol, and the standard state for water is one in which the vapor pressure of water is 3.7 times as great as that of pure water. The fact that these standard states are unattainable in practice causes no real difficulty; the free energy change in any actual process, such as that of mixing two components to form a solution of given composition, may be calculated in terms of either of the two standard states we have defined—or indeed in terms of any other standard state that may be found convenient in a given problem. It is of the utmost importance, however, once a standard state for a given component has been adopted, that the convention chosen be rigorously adhered to in the subsequent calculations.

Consider for instance the free energy change which occurs when 9 moles of water (n_1) are mixed with 1 mole of propanol (n_2) to form a solution of mole fraction $N_2 = 0.1$. If we choose pure water and pure propanol as the standard states, then the free energy change at 298.1° K, taking $R = 1.987 \text{ cal deg}^{-1} \text{ mole}^{-1}$, is

$$\begin{aligned}
 \Delta F &= n_1(\mu_1 - \mu_1^\circ) + n_2(\mu_2 - \mu_2^\circ) & (75) \\
 &= 9RT \ln a_1 + RT \ln a_2 \\
 &= RT(9 \ln N_1 + \ln N_2 + 9 \ln f_1 + \ln f_2) \\
 &= 1365(9 \log N_1 + \log N_2 + 9 \log f_1 + \log f_2) \\
 &= 1365(-0.414 - 1.00 + 0.225 + 0.772) \\
 &= 1365(-1.414 + 0.997) = -1930 + 1360 = -570 \text{ cal}
 \end{aligned}$$

The ideal free energy of mixing is -1930 calories, given by the terms in $\log N_1$ and $\log N_2$; the actual free energy of mixing is numerically much less, because the term due to the activity coefficients is opposite in sign to the ideal term.

If we choose to employ the second convention concerning activity coefficients, we must make allowance for the fact that the activity of pure water (which we may call a_1^*) is 0.270 by this convention, being numerically equal to the corresponding activity coefficient, $f_1'^* = p_1^\circ/k_1$ for pure water. Likewise for pure propanol $a_2^* = f_2'^* = p_2^\circ/k_2 = 0.0694$. The free energy of mixing 9 moles of water with 1 mole of propanol is

$$\begin{aligned}
 \Delta F &= 9RT \ln (a_1/a_1^*) + RT \ln (a_2/a_2^*) \\
 &= RT[9 \ln N_1 + \ln N_2 + 9 \ln (f_1'/f_1'^*) + \ln (f_2'/f_2'^*)]
 \end{aligned}$$

In terms of the definitions of $f_1'^*$ and $f_2'^*$ given above, however, and those of f_1' and f_2' given in (73) and (74), it is apparent that $f_1'/f_1'^* = f_1$, and $f_2'/f_2'^* = f_2$, so that this equation for ΔF becomes identical with the other, as of course it should. Obviously, for the solution of this particular problem, the first convention concerning standard states is much more convenient to use in practice than the second.

If we are dealing with dilute solutions, however, it is often most convenient to choose the activity coefficient for the solute by the second convention, so that its activity coefficient becomes equal to unity as its mole fraction approaches zero. At the same time we may use the first convention for the solvent, setting its activity equal to unity in the pure state. In the dilute aqueous solutions, in which many biochemical components are studied experimentally, the mole fraction of water is often so nearly unity that we are justified in applying Raoult's law and setting the activity of the water equal to its mole fraction, if we do not have a direct measurement of its vapor pressure in the particular system under study. Indeed, except in calculations of high precision, it is often sufficient to set $a_1 = 1$ for the activity of water in such systems.

Very commonly it is convenient to use activity coefficients for the solute which are defined by the ratio of the activity of the solute to its molality m (moles of solute per kilogram of solvent), or to its molar concentration c (moles of solute per liter of solution), in either case employing the convention that a_i/m_i or a_i/c_i approaches unity as m_i or c_i , respectively, approaches zero. In dilute aqueous solutions, when c_i or m_i is (say) below 0.1, and N_i is therefore below 0.002, the numerical value of the activity coefficient is so nearly the same on all three concentration scales that we may often ignore the difference in many biochemical systems, in which the measurements are not of very high precision. At higher concentrations, however, c_i , m_i , and N_i do not bear the same relation to one another that they do at very high dilutions, and the choice of the unit used to denote the solute concentrations affects the numerical value of the activity coefficient. The relations between the activity coefficients, expressed in terms of the different concentration scales, are succinctly stated by Glasstone (1947, p. 355). In any case the reader should note that the ratio of the *activities* of a given solute in two different solutions—provided both are expressed in terms of the same standard state—must be independent of any convention, since this ratio is related to the difference in the chemical potential of the solute in the two states by the equation

$$(\mu_i)_1 - (\mu_i)_2 = RT \ln [(a_i)_1/(a_i)_2] \quad (76)$$

and this difference in chemical potential is independent of the arbitrary choice of concentration scales.

Activity and Chemical Potential of a Nonvolatile Solute from the Vapor Pressure of the Solvent

We may calculate the activity of a nonvolatile solute, such as an amino acid or sugar, if the vapor pressure of the solvent is known over the whole range of solute concentration from zero to any upper limit in which we are interested. We consider aqueous solutions specifically, and choose to express the concentration of solute on the molality scale; hence the solution, if it is a two-component system, contains 1 kg or 55.51 moles of water (component 1) and m moles of solute (component 2). The activity of the solute is $a_2 = f_2 m$. The Gibbs-Duhem equation (67) then becomes

$$\begin{aligned} m \, d \ln a_2 &= m \, d \ln m + m \, d \ln f_2 \\ &= -55.51 \, d \ln a_1 = -55.51 \, d \ln (p_1/p_1^\circ) \quad (77) \end{aligned}$$

To facilitate integration of (77), we introduce a quantity ϕ , commonly called the molal osmotic coefficient of the solute, defined by the equation

$$\phi m = -55.51 \ln a_1 \quad (78)$$

To determine ϕ we need to know only the vapor pressure of the solvent and the molality of the solute, which are both readily accessible to experimental measurement. Since Raoult's law applies to component 1 when its activity approaches unity ($N_1 \rightarrow 1$), ϕ must approach unity as m approaches zero.⁸ Differentiating (78), we obtain with the aid of (77)

$$d(\phi m) = \phi \, dm + m \, d\phi = m \, d \ln m + m \, d \ln f_2$$

Dividing by m and rearranging,

$$(\phi - 1) \, d \ln m + d\phi = d \ln f_2 \quad (79)$$

Integration of (79), since $\phi \rightarrow 1$ as $m \rightarrow 0$, gives

$$\ln f_2 = \int_0^m (\phi - 1) \, d \ln m + (\phi - 1) = \int_0^m \frac{(\phi - 1)}{m} \, dm + (\phi - 1) \quad (80)$$

The integration may be carried out either graphically or analytically. The integral in (80) may be evaluated graphically by plotting $(\phi - 1)/m$ against m . However, since both the numerator and the denominator of this expression tend to zero as m vanishes, their ratio at

⁸ As N_1 approaches 1, a_1 approaches $N_1 = 1 - N_2$. A series expansion gives

$$\ln (1 - N_2) = -N_2 - N_2^2/2 - N_2^3/3 \dots$$

When N_2 is very small, it becomes equal to $m/55.51$ in the limit. It is then readily shown from (78) that ϕ must approach unity.

low m values becomes uncertain unless the experimental data are of very high precision.

If the activity of water has been determined with high precision in solutions of one nonvolatile solute, which is taken as a reference standard, then it may readily be determined in solutions of other such solutes, by what is known as the isopiestic vapor pressure method. Two open containers are set up in a closed vessel, one containing a solution of the reference standard substance, the other a solution of the substance being investigated. If the vapor pressures of the two solutions are different, water will evaporate from the solution with the higher vapor pressure into the surrounding closed space, and will condense into the solution with the lower vapor pressure. When equilibrium is attained, the activity of water in the two solutions must be equal. The two solutions are then analyzed; if the molality of the standard is m_s and that of the unknown is m_x , then the activity of the water in both solutions is known from the previous studies on the standard, and the osmotic coefficient of substance x at molality m_x is immediately calculated from (78). Repetition of the experiment for a wide range of such isopiestic solutions giving paired values of m_s and m_x provides the data necessary for the calculation of the activity coefficient of x by equation (80). To accelerate the attainment of equilibrium it is desirable to remove most of the air from the vessel enclosing the containers, and to maintain thermal equilibrium the containers are made of silver, or some other highly conducting material, and placed on a block of copper. The reference substances most commonly employed have been sucrose and potassium chloride. Extensive studies on the activity coefficients of amino acids by this method have been carried out by P. K. Smith and E. R. B. Smith (1937, 1940). Some of the data are shown in Table II.

The isopiestic vapor pressure method is excellent for values of m above 0.1 or thereabouts, but is subject to uncertainty in more dilute solutions. At high dilutions the determination of the activity of the solute from the freezing point depression of the solvent is the method of higher precision. Since it is treated in some detail in all the principal textbooks of chemical thermodynamics listed at the end of this chapter, and since an accurate formulation of this method requires a careful and detailed discussion, we shall not treat it here. We may note, however, that, since this method determines the activity of the solvent (and, by calculation, of the solute) at the freezing point of the solution, additional calculations of the temperature coefficient of the activity must be carried out if activity coefficients are to be determined at higher temperatures, say at 25° or 37°. This involves knowledge of the heat changes accompanying dilution of the solution under study. Particularly accurate activity coefficient values

TABLE II
ACTIVITY COEFFICIENTS OF CERTAIN AMINO ACIDS AND PEPTIDES IN AQUEOUS SOLUTION AT 25°
[From the measurements of Smith and Smith (1937, 1940)]

Moles per 1000 g of water m	Gly- cine* - log f	Ala- nine - log f	Sarco- sine - log f	Serine - log f	Threo- nine - log f	Valine - log f	Betaine - log f	Proline - log f	Hy- droxy- proline - log f	β -Ala- nine - log f	ϵ - caproic acid - log f	Glycyl- glycine - log f	Glycyl- alanine - log f	Tri- glycine - log f
0.2	0.0173	-0.002	-0.002	0.016	0.005	-0.013	-0.030	-0.008	0.000	0.003	0.013	0.040	0.029	0.070
0.3	0.0252	-0.003	-0.003	0.025	0.007	-0.019	-0.044	-0.012	0.000			0.056	0.040	0.095
0.4	0.0326	-0.004				-0.025								
0.5	0.0395	-0.005	-0.005	0.042	0.011	-0.032	-0.074	-0.020	-0.001	0.005	0.022	0.082	0.054	
0.7	0.0521	-0.009	-0.009		0.015		-0.104	-0.029	-0.001	0.006	0.024	0.102	0.061	
0.8	0.0578	-0.008												
1.0	0.0684	-0.010	-0.014		0.018		-0.147	-0.040	-0.003	0.005	0.026	0.128	0.068	
1.2	0.0780	-0.012	-0.018		0.020		-0.176		-0.004	0.004	0.016	0.141	0.073	
1.5	0.0903	-0.015	-0.025		0.022		-0.219	-0.060	-0.006	0.000	-0.001	0.157		
1.7	0.0976	-0.017										0.164		
1.9		-0.020												
2.0	0.107		-0.038		0.025		-0.289	-0.081	-0.011	-0.007	-0.030			

All activity coefficients were determined by the isopiestic vapor pressure method (see text) and are expressed as $f = a/m$, where m is molality (moles per 1000 g of water). The values for glycine are slightly revised from those of Smith and Smith in the light of other measurements; see Cohn in Cohn and Edsall (1943, Chapter 10).

* The value of $-\log f$ for glycine at $m = 2.5$ is 0.121; the value at $m = 3.3$ (saturated solution) is 0.137.

by the freezing point method have been obtained by Scatchard and his associates (Scatchard and Prentiss, 1934; Scatchard, 1936).

Equilibrium between Phases; Activities in Relation to Distribution Coefficients and Solubilities

When two or more different phases are in equilibrium, the chemical potential of each component must be the same throughout all the phases—at least in all phases in which the component in question is present in appreciable amounts. This qualification is necessary, since in a solubility determination, for instance, the solid phase may consist of only one component. In that case we cannot make any quantitative statement concerning the potentials of the other components in that phase.

Thus in the distribution of a solute (component i) between two liquid phases at equilibrium, we know that its chemical potential must be the same in both. If we denote one phase by the subscript 1, and the other by the subscript 2, then

$$(\mu_i)_1 = (\mu_i)_2 = RT \ln (a_i)_1 + (\mu_i^\circ)_1 = RT \ln (a_i)_2 + (\mu_i^\circ)_2 \quad (81)$$

Equation (81) indicates that the standard states chosen for component i may be different in the two phases, 1 and 2. If the same standard state is chosen for both phases, then of course $(a_i)_1 = (a_i)_2$ at equilibrium. The equality of the chemical potentials for component i in the two phases, however, must hold regardless of the choice of standard state.

We may, for instance, study the distribution of an organic molecule between water (component 1) and benzene (component 2), which are almost completely immiscible. If we wish, for instance, to determine the effect of added salt (component 3) on the solubility of the organic solute (component 4), then we may make up a series of benzene-water systems in which the aqueous phase contains various salt concentrations, from zero up. The organic solute may be added to each of these systems to the extent that, after equilibration, its concentration in the benzene phase is the same in all the systems. Since the benzene phase is essentially pure benzene in all cases, except for component 4, this means that a_4 in benzene has been adjusted to the same value in all cases. Since the aqueous solutions are all in equilibrium with the same benzene phase, a_4 must therefore be the same in all aqueous solutions. Hence, since by definition, for any component,

$$a_i = f_i c_i \quad (82)$$

the activity coefficient of component 4 in the aqueous phase varies inversely as its concentration in that phase at equilibrium. A system of this sort has been used by Ph. Gross to study the effect of various salts

on the activity coefficients of acetone and hydrocyanic acid. The results are described in Chapter 5.

The determination of activity coefficients from solubility measurements is based on exactly the same principle. The potential of any component present in the solid phase must be the same as it is in any saturated solution in equilibrium with that phase. Consequently, if the composition of the solvent is varied by adding some other component, the chemical potential of the solute present in the solid phase must remain fixed in the saturated solution. Hence the activity of the solute is also fixed, on the assumption that we refer all solubility measurements to the same standard state. As the solubility increases, the activity coefficient decreases, and vice versa. Thus equation (82), or an analogous equation expressed in terms of the mole fraction of solute, applies to this case also. An exam-

TABLE III
SOLUBILITY OF ASPARAGINE IN SODIUM CHLORIDE SOLUTIONS AT 25°

NaCl (M/l)	ρ	Solubility, c_2 (M/l)	$-\log \gamma = \log (c_2/c_2^\circ)$	N_2	$-\log f = \log (N_2/N_2^\circ)$
0.0	1.00714	0.184		0.00336	0.0
0.25	1.01821	0.203	0.043	0.00371	0.043
0.50	1.02848	0.216	0.070	0.00396	0.071
1.00	1.04881	0.241	0.117	0.00442	0.119
2.00	1.08782	0.290	0.198	0.00536	0.203

From data in Cohn and Edsall (1943), Chapter 11.

Asparagine is denoted as component 2. Its activity coefficient is expressed both as $-\log \gamma = \log (c_2/c_2^\circ)$ and as $-\log f = \log (N_2/N_2^\circ)$ where the circle superscript denotes solubility of asparagine at zero sodium chloride concentration. The symbol ρ denotes the density of the solution.

ple of the determination of activity coefficient from solubility measurements is given in Table III, taken from the work of Cohn and McMeekin (see Cohn and Edsall, 1943, p. 241) on the solubility of asparagine in sodium chloride solutions. It is obvious from these data that, as the concentration of sodium chloride increases, the solubility of asparagine also increases, and the activity coefficient decreases. $\log f$, at low salt concentrations, is approximately a linear function of sodium chloride concentration. We take the activity coefficient as unity in the solution containing no sodium chloride in Table III, noting that a slightly different value would probably be obtained if we chose asparagine at infinite dilution in water as the standard state. The significance of such data, in terms of electrostatic interactions between the ions of sodium chloride and the dipolar ion asparagine, is considered in Chapter 5, where a number of further examples of such interactions are given.

One warning should be given with regard to the use of solubility measurements in determining activity coefficients. It is essential that in all the solutions under comparison the composition of the solid phase in equilibrium with the solution must be the same. Sometimes the addition of new components to the solvent in considerable amount results in the formation of a new solid phase. If this occurs, the chemical potential of the components of the solid phase must inevitably be changed by the change in crystal composition, and it follows that the activity of any such components in the saturated solution at equilibrium will no longer be the same as before. In solubility studies on most simple inorganic or organic compounds this complication does not arise, even though the composition of the solution may vary widely. Changes in the composition of the solid phase, however, do sometimes occur, and the possibility of their occurrence must always be considered. A further discussion of this point is given in Chapter 5, p. 320.

Activity of Strong Electrolytes

The definition of activities and activity coefficients for strong electrolytes involves some special considerations. It is desirable to define these quantities so that the activity of each ionic species becomes equal to its concentration at infinite dilution; the concentration may be defined in terms of molar concentration, molality, or mole fraction, as with non-electrolytes. Generally one of the two former units is employed; most physical chemists have expressed concentrations in terms of molalities (see, for instance, Robinson and Stokes, 1955).

The simplest case is that of an electrolyte, such as NaCl or MgSO_4 , which yields one anion and one cation. If we denote the electrolyte (anion plus cation) as component 2, its activity, a_2 , is defined by the usual relation:

$$\mu_2 - \mu_2^\circ = RT \ln a_2 \quad (83)$$

The potential μ_2 may be taken as equal to the sum of the potentials of the constituent ions, μ_+ and μ_- :

$$\begin{aligned} \mu_2 &= \mu_+ + \mu_- \\ \mu_2^\circ &= \mu_+^\circ + \mu_-^\circ \end{aligned} \quad (84)$$

Hence we obtain

$$RT \ln a_+ + RT \ln a_- = RT \ln a_2 = 2RT \ln a_{\pm} \quad (85)$$

or

$$a_+ a_- = a_2 = (a_{\pm})^2$$

where a_{\pm} is a mean ionic activity coefficient, defined by this equation.

The activity coefficient of each ion may be expressed in terms of its activity coefficient, f_+ or f_- , and its concentration, m_+ or m_- :

$$f_+ = a_+/m_+; \quad f_- = a_-/m_- \quad (86)$$

and the mean ionic activity coefficient, f_+ , is defined by

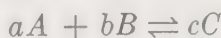
$$f_{\pm}^2 = f_+f_- = (a_+a_-)/(m_+m_-) = (a_{\pm})^2/(m_{\pm})^2 \quad (87)$$

where $m_{\pm}^2 = m_+m_- = m^2$, m being the molality of the electrolyte as a whole. This definition is readily extended to electrolytes giving more than two ions; the formulas involved, and some of their uses, are given in the discussion of the salting-out effect and the Debye-Hückel theory in Chapter 5.

All the methods previously referred to for determining activities and activity coefficients of involatile nonelectrolytes may also be applied to electrolytes. In addition, there is a powerful method, applicable to electrolytes only—the determination of the electromotive force of a galvanic cell, operated under reversible conditions, in which the electrolyte under study is formed in, or disappears from, the solution between the electrodes when current flows. The use of a certain class of such cells for the study of acid-base equilibria is described in Chapter 8; more general discussions of the thermodynamics of galvanic cells are given in some of the books listed at the end of this chapter.

Mass Law

Now that we have developed the concepts of chemical potential and activity, we are in a position to discuss the question of equilibrium in a chemical reaction. To make matters simple and concrete, we shall suppose the reaction to be one in which a moles of component A combine with b moles of component B to give c moles of component C :



The treatment can, of course, easily be generalized. The criterion of equilibrium is that for any infinitesimal variation of the system associated with a reaction either to the right or the left there be no change of free energy, pressure and temperature remaining constant. For such a variation, involving $a \, dn$ moles of A and $b \, dn$ moles of B

$$dF = (c\mu_C - a\mu_A - b\mu_B)dn$$

Consequently the condition of equilibrium ($dF = 0$) becomes

$$c\mu_C - a\mu_A - b\mu_B = 0 \quad (88)$$

This fundamental equation assumes a more familiar form if we introduce

activities in place of chemical potentials. In order to do this, we subtract from both sides of it the expression

$$c\mu_C^\circ - a\mu_A^\circ - b\mu_B^\circ = \Delta F^\circ \quad (89)$$

In this expression, the superscript $^\circ$ indicates that each chemical potential corresponds to the standard state, and ΔF° is the free energy change of the reaction when all reactants and products are in their standard states. For convenience, we shall refer to this as the standard-state reaction. On subtracting (89) from (88) we obtain

$$c(\mu_C - \mu_C^\circ) - a(\mu_A - \mu_A^\circ) - b(\mu_B - \mu_B^\circ) = -\Delta F^\circ$$

Since for any component the activity, a , is given by

$$\mu_i - \mu_i^\circ = RT \ln a_i$$

this last equation yields

$$\log \frac{a_C^c}{a_A^a a_B^b} = -\frac{\Delta F^\circ}{RT} \quad (90)$$

or

$$\frac{a_C^c}{a_A^a a_B^b} = e^{-\Delta F^\circ/RT}$$

At any given temperature and pressure, ΔF° and, therefore, $e^{-\Delta F^\circ/RT}$ are constants. Consequently equation (90), except for the substitution of activities for concentrations, is the same as the mass law in its familiar form. It is in fact the valid form of that law, the ordinary form, involving concentrations, being an approximation, accurate only to the extent to which concentrations are proportional to activities.

Equation (90) shows how the true equilibrium constant, K , of a chemical reaction is related to ΔF° :

$$K = e^{-\Delta F^\circ/RT}$$

or

$$\Delta F^\circ = -RT \ln K \quad (91)$$

Here ΔF° , which is a change of free energy, gives the total work, over and above pressure-volume work, which can be obtained from the standard-state reaction when it occurs reversibly. Consequently, if we have any means of making the reaction occur reversibly, this will provide a basis of determining the equilibrium constant. Such a means is often found in a galvanic cell. It is for this reason that galvanic cells are of such importance in the study of equilibria and chemical potentials. The use of certain galvanic cells for these purposes is discussed in Chapter 8.

The term ΔF° , like the chemical potentials μ° to which it is related, is a function of pressure and temperature, and it is of importance to derive

expressions for the change of ΔF° , and consequently of K , with these variables. Since, as we have shown, for any component, i ,

$$\left(\frac{\partial \mu_i}{\partial P}\right)_T = \bar{V}_i$$

it follows at once from (89) that for the reaction we have been discussing

$$\left(\frac{\partial \Delta F^\circ}{\partial P}\right)_T = c\bar{V}_{C^\circ} - a\bar{V}_{A^\circ} - b\bar{V}_{B^\circ}$$

In accordance with the nature of partial molal quantities, however, the right-hand member of this equation gives the change of volume of the system accompanying the standard-state reaction. If we denote this by ΔV° , we can write, therefore,

$$\left(\frac{\partial \Delta F^\circ}{\partial P}\right)_T = \Delta V^\circ \quad (92)$$

Combination of (91) and (92) yields immediately

$$RT \left(\frac{\partial \ln K}{\partial P}\right)_T = -\Delta V^\circ \quad (93)$$

Equation (93) shows that if a reaction involves an increase of volume it will be suppressed by raising the pressure. For gas reactions, where volume changes may be large, the effect of pressure is often great. For reactions in condensed phases, it is in general much less.

The derivative of $\ln K$ with respect to temperature is obtained at once from (30.2). It is

$$\frac{\partial}{\partial T} \left(\frac{\Delta F^\circ}{T}\right)_P = -R \left(\frac{\partial \ln K}{\partial T}\right)_P = -\frac{\Delta H^\circ}{T^2} = -\frac{\Delta S^\circ}{T}$$

or

$$RT^2 \left(\frac{\partial \ln K}{\partial T}\right)_P = \Delta H^\circ \quad (94)$$

This is generally known as van't Hoff's equation. It may also be written

$$\left[\frac{\partial \ln K}{\partial (1/T)}\right]_P = -\frac{\Delta H^\circ}{R} \quad (95)$$

If ΔH° can be taken as constant over a temperature range from T_1 to T_2 , (95) may be immediately integrated to give

$$\ln \left(\frac{K_2}{K_1}\right) = \frac{\Delta H^\circ}{R} \left(\frac{1}{T_1} - \frac{1}{T_2}\right) \quad (96)$$

If we convert to denary logarithms, and express ΔH° in calories per mole, this becomes

$$\log \left(\frac{K_2}{K_1} \right) = \frac{\Delta H^\circ}{4.576} \left(\frac{1}{T_1} - \frac{1}{T_2} \right) \quad (97)$$

We may use this equation either to calculate the heat of reaction if the equilibrium constant is known at two or more different temperatures, or to calculate the change of equilibrium constant with temperature if the heat of reaction is known. If $4.576 \log K$ is plotted against $1/T$, the resulting curve should be a straight line if ΔH° is independent of temperature, and the slope gives ΔH° directly. If the curve is not linear, the slope of the tangent to it, at any given value of $1/T$, gives ΔH° for the corresponding value of T . In many reactions ΔH° does vary considerably with temperature, and the integration of (95) requires the formulation of ΔH° as a function of temperature and the integration of the resulting expression. We note that, from (37.8), the temperature derivative of ΔH is

$$\left(\frac{\partial \Delta H}{\partial T} \right)_p = \Delta C_p$$

Hence for a reaction involving a large change in the heat capacity of the system, the heat of reaction must vary rapidly with temperature. This is true, for instance, for the ionization of a neutral acid (HA) such as acetic acid, according to the reaction

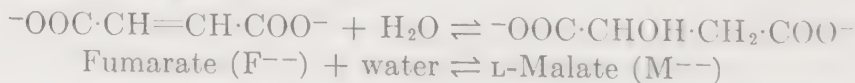


Reactions of this type characteristically have ΔC_p values of the order of -40 to -50 cal deg $^{-1}$ mole $^{-1}$ (see Chapter 8, Table III). The decrease in the heat capacity is due to the formation of two ions from two uncharged molecules and the orientation and compression of the surrounding water around the ions (see p. 173).

We now offer three illustrations of the application of the law of mass action in some important biochemical systems.

THE MALATE-FUMARATE EQUILIBRIUM

A reaction of major importance in the well-known tricarboxylic cycle of intermediary metabolism is the interconversion of fumarate and L-malate according to the reaction



This reaction readily comes to equilibrium in the presence of the enzyme fumarase, although in the absence of a catalyst equilibrium is attained only with extreme slowness, if at all. Since the reaction is generally car-

ried out in a very dilute aqueous solution, the activity of the water may be considered as practically unity, and the equilibrium constant may be written

$$K_{\text{MF}} = (a_{\text{M}^{--}})/(a_{\text{F}^{--}}) \cong (M^{--})/(F^{--}) \quad (98)$$

That is, we make the approximate assumption that the activity of the malate and fumarate ions may be set equal to their molar concentrations. Among the most recent measurements of the equilibrium are those of Krebs (1953). Krebs worked at 25° at a total concentration of malate plus fumarate of 0.05 *M*, in the presence of a fumarase preparation. The

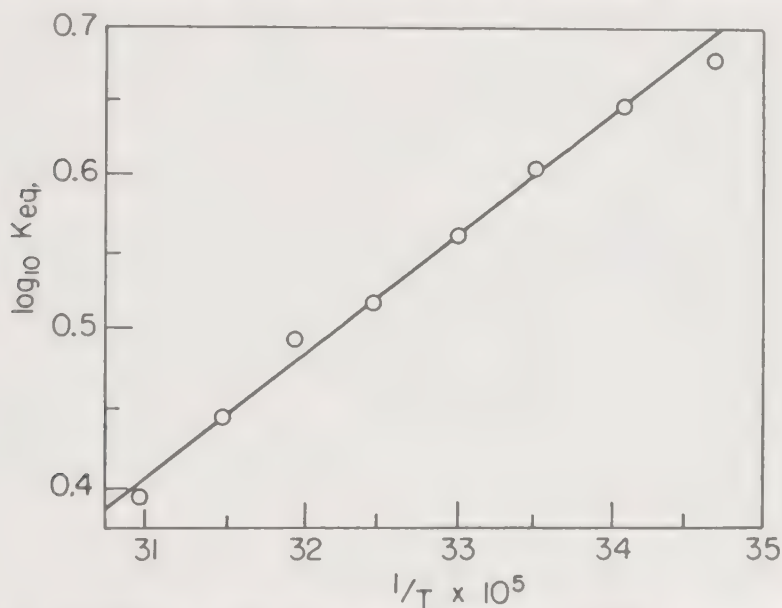


FIG. 5. Temperature dependence of the malate-fumarate equilibrium. (From Scott and Powell, 1948.)

same equilibrium point was attained in a series of measurements, and the value K_{MF} was found to be, at pH 7.4,

$$K_{\text{MF}} = 4.42 \pm 0.10 \quad (T = 298.1^\circ \text{K}, \text{pH } 7.4)$$

This gives for the standard free energy of the reaction

$$\Delta F^\circ = -RT \ln K = -880 \text{ cal mole}^{-1} \quad (99)$$

The value of K_{MF} decreases with rising temperature, showing that heat is evolved when the reaction proceeds from left to right. The temperature coefficient of the equilibrium has been determined by Krebs *et al.* (1940) and by Scott and Powell (1948), with results in reasonably satisfactory agreement. The data of the latter authors are shown in Fig. 5. It is seen that $\log K$ is very nearly a linear function of $1/T$, and the value of ΔH° , from equation (97), is $-3560 \text{ cal mole}^{-1}$. This should also be close to the

actual value of ΔH in the solution under the experimental conditions. We note that the value of K_{MF} at 25°, according to Scott and Powell, is only 4.07, significantly lower than the value of 4.42 reported by Krebs. We take the latter as giving the most reliable value for ΔF° at this temperature, but employ the values given by the former authors for calculating ΔH° .

The standard entropy change in the reaction at 25° (298.1° K) is

$$\Delta S^\circ = (\Delta H^\circ - \Delta F^\circ)/T = (-3560 + 880)/298.1$$

$$= -9.0 \text{ cal deg}^{-1} \text{ mole}^{-1} \quad (100)$$

In any reaction involving acids and bases, the position of the equilibrium is likely to be affected by changes of the hydrogen ion activity, a_H , usually expressed as its negative logarithm, the pH . Although a full discussion of acid-base equilibria and pH is deferred to Chapters 8 and 9, the subject is of such importance for the formulation of biochemical equilibria that we anticipate some of the conclusions of that chapter to consider the effect of pH on the fumarate-malate equilibrium. The reader who is unfamiliar with the concepts employed in the following discussion should defer his study of it until he has become familiar with the contents of Chapter 8 and the opening section of Chapter 9. Here we have closely followed the treatment of Krebs (1953).

Fumaric and malic acids are both dibasic acids; the former can exist in the forms H_2F , HF^- , and F^{--} , the latter in the forms H_2M , HM^- , and M^{--} . The ionization constants of H_2F may be written

$$K_{1F} = (H^+)(HF^-)/(H_2F) = 9.6 \times 10^{-4};$$

$$K_{2F} = (H^+)(F^{--})/(HF^-) = 4.0 \times 10^{-5} \quad (101)$$

with analogous equations for the ionization constants of malic acid, for which:

$$K_{1M} = 3.3 \times 10^{-4}; K_{2M} = 7.7 \times 10^{-6} \quad (102)$$

The quantity measured by an analytical determination for fumaric acid in solution is the total concentration, T_F , of fumaric acid in all its forms

$$T_F = (H_2F) + (HF^-) + (F^{--})$$

$$= (F^{--})[K_{1F}K_{2F} + (H^+)K_{1F} + (H^+)^2]/K_{1F}K_{2F} \quad (103)$$

with an exactly analogous expression for the total malate, T_M , as a function of K_{1M} , K_{2M} , and (H^+) . Taking the ratio of these two expressions we obtain an equation for the ratio $(M^{--})/(F^{--})$:

$$\frac{(M^{--})}{(F^{--})} = \frac{T_M}{T_F} \times \frac{K_{1M}K_{2M}[K_{1F}K_{2F} + K_{1F}(H^+) + (H^+)^2]}{K_{1F}K_{2F}[K_{1M}K_{2M} + K_{1M}(H^+) + (H^+)^2]} = \frac{T_M}{T_F} Q \quad (104)$$

At pH 7.4 and above, (H^+) is so small compared with the K 's that Q is practically unity, and $(M^{--})/(F^{--})$ is virtually identical with T_M/T_F .

These are the conditions under which the value of ΔF° in (99) was obtained. With decreasing pH [increasing (H^+)], however, Q deviates more and more from unity, and the value of T_M/T_F correspondingly alters. Calculated values of this quantity are given in Table IV for different

TABLE IV

THE RATIO (TOTAL MALATE)/(TOTAL FUMARATE) AS A FUNCTION OF pH AT 25°

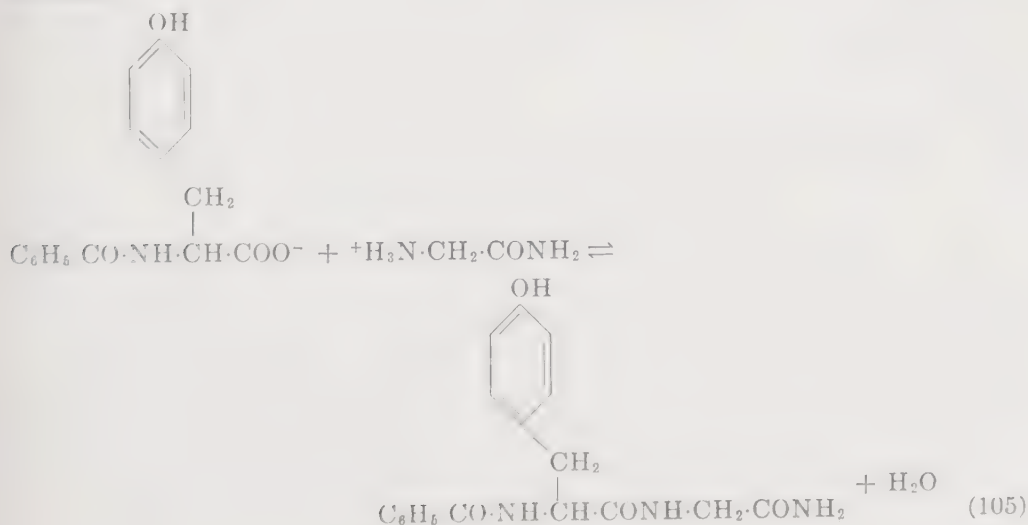
pH	T_M/T_F	pH	T_M/T_F
7.4	4.42	4.0	22.4
7.0	4.46	3.0	44.6
6.0	4.86	2.0	62.0
5.0	8.26	1.0	66.4

From Krebs (1953).

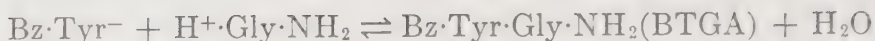
values of $pH = -\log (H^+)$. Experimentally it is impossible to follow the reaction much below pH 5, since the catalytic activity of the enzyme becomes very weak in acid solution. Krebs obtained one measurement at pH 4.85, which gave a value of $T_M/T_F = 8.1$, a value somewhat lower than that to be expected from Table IV, but not far off. The deviation was not surprising, as the values of K_{1F} , K_{2F} , K_{1M} , and K_{2M} employed were determined under somewhat different conditions from those in the experimental solutions.

PEPTIDE BOND SYNTHESIS IN THE REACTION BETWEEN BENZOYL-L-TYROSINE AND GLYCINAMIDE

These two substances react in the presence of the enzyme chymotrypsin to form benzoyl-L-tyrosyl-glycinamide. The reaction may be written:



or, in brief,



The reaction was studied by Dobry *et al.* (1952), who used glycynamide labeled with a known excess of the stable isotope N^{15} . The system, at 25° and $\text{pH } 7.90$, was incubated 6 hours in the presence of the enzyme; this time was shown by other experiments to be sufficient for attainment of equilibrium. The reaction was then stopped by pouring the solution into a large excess of absolute ethanol, which contained a considerable amount of nonisotopic BTGA. Then BTGA was isolated from the mixture, carefully recrystallized to free it from impurities, and the N^{15} content of the product determined. From this figure, and the known N^{15} content of the original glycynamide, the amount of BTGA formed in the reaction was immediately calculated. With the amounts of the reactants originally added to the system known, it was then calculated that the concentrations at equilibrium were: benzoyl-L-tyrosine, $0.025 \text{ } M$; glycynamide, $0.05 \text{ } M$; BTGA, $0.00032 \text{ } M$. We cannot calculate the equilibrium constant for reaction (105), however, without taking account of one further fact. The *total* concentration of glycynamide is $0.05 \text{ } M$; this is the sum of $(^+\text{H}_3\text{N}\cdot\text{CH}_2\cdot\text{CONH}_2)$ and $(\text{H}_2\text{N}\cdot\text{CH}_2\cdot\text{CONH}_2)$; but only the concentration of the former enters into reaction (105). Independent measurements of the acid-base equilibrium (Chapter 8) in glycynamide have established the constant K_A :

$$K_A = \frac{(\text{H}^+)(\text{H}_2\text{N}\cdot\text{CH}_2\cdot\text{CONH}_2)}{(^+\text{H}_3\text{N}\cdot\text{CH}_2\cdot\text{CONH}_2)} = 10^{-7.93} \quad (106)$$

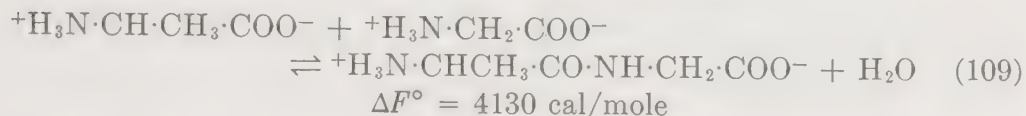
The experiment was carried out at $\text{pH } 7.90$ [$(\text{H}^+) = 10^{-7.90}$]. Thus at this pH it may be calculated from (106) that $(^+\text{H}_3\text{N}\cdot\text{CH}_2\cdot\text{CONH}_2)$ represents 52% of the total glycynamide present: its concentration is therefore $0.05 \times 0.52 = 0.026$. Hence the equilibrium constant⁹ for the formation of BTGA is

$$K = \frac{(\text{BTGA})}{(\text{Bz}\cdot\text{Tyr}^-)(^+\text{HGly}\cdot\text{NH}_2)} = \frac{0.00032}{0.025 \times 0.026} = 0.492 \text{ liter mole}^{-1} \quad (107)$$

$$\Delta F^\circ = -RT \ln K = 420 \text{ cal mole}^{-1} \quad (108)$$

⁹ The reader may object that we have assumed, without proof, that the benzoyl-L-tyrosine ($\text{R}\cdot\text{COO}^-$) is entirely in the ionized form. If some were in the form $\text{R}\cdot\text{COOH}$, the total benzoyltyrosine would be greater than the concentration of the anionic form, and we should have to correct for this as we have done for the glycynamide. The value of $\text{p}K_A = -\log K_A$ for the carboxyl group in benzoyltyrosine is near 3.7, however, and this means that less than 0.5% is in the un-ionized form at any pH above 6 (less than 0.01% at $\text{pH } 7.9$). We are therefore quite justified in neglecting this complication

The estimated experimental uncertainty was ± 50 cal mole⁻¹. The synthesis of BTGA is an endothermic process; $\Delta H^\circ = 1550 \pm 100$ cal/mole. The value of ΔF° in (108) is relatively small; it indicates an appreciable tendency to synthesis of the peptide bond at equilibrium. In contrast it has been calculated (Borsook, 1953) that the standard free energy change for the synthesis of alanylglycine from alanine and glycine is very much larger. At 37.5° Borsook calculates as follows:



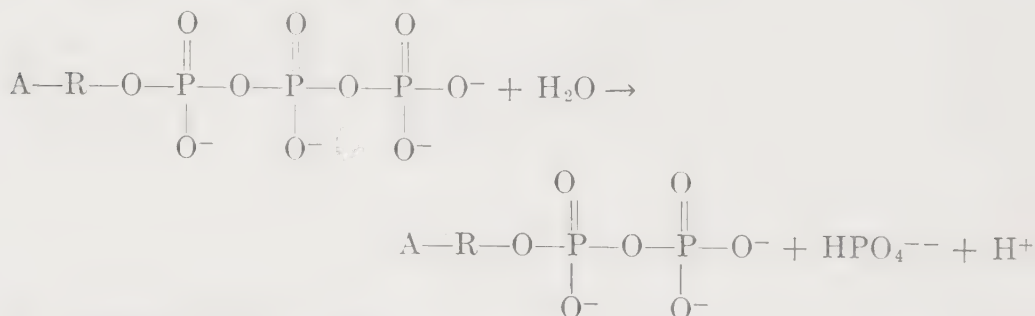
A similar but slightly lower value ($\Delta F^\circ = 3590$ cal/mole) is calculated for the synthesis of glycylglycine from two molecules of glycine. Reactions with so large a positive standard free energy change do not take place to any significant extent. The most striking difference between (105) and (109) is that in the former the product of synthesis (BTGA) contains no charged groups, whereas in (109) it contains both an anionic and a cationic group. These charges set up electrostatic fields which tend to hinder the approach of the reacting amino and carboxyl groups in (109) and thereby make synthesis of the peptide linkage more difficult. Since the synthesis of a protein molecule involves the synthesis of a great number of peptide linkages, the importance of these free energy considerations for protein biosynthesis is obvious.

Such a reaction as (105) or (109), with a numerically positive ΔF° value, will proceed from left to right to a limited extent in the presence of an enzyme to catalyze it; but, especially if ΔF° is as large and positive as for (109), the extent of synthesis will be very small. From the nature of the equilibrium constants involved, it is obvious that increasing the total concentration of reactants increases the percentage of reactant molecules which have joined to form this peptide bond at equilibrium. In biological fluids and cells, however, the concentrations of free amino acids and peptides are extremely low, and the syntheses actually achieved cannot be explained merely by postulating high concentrations of these materials in local regions of the cell. Other possible mechanisms must be considered. One method of making a reaction with a large positive value of ΔF° proceed from left to right is to remove the product of the reaction as fast as it is formed. This may occur if the product is very insoluble and crystallizes out of the liquid phase. It may also occur if the reaction product enters immediately into another reaction which consumes it. In the complex reaction sequences which proceed in biochemical systems, this of course frequently occurs. For a *total* process, however, it is impossible to escape the rigorous requirement that the *total* free energy change, at

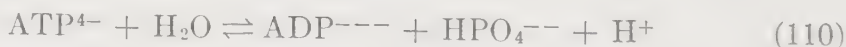
constant pressure and temperature, must be negative if the process is to go. Such spontaneous processes (ΔF negative) have been termed exergonic by C. D. Coryell, in analogy to the term exothermic for processes in which ΔH is negative. A process which taken alone would be endergonic (ΔF positive), and therefore thermodynamically forbidden, may nevertheless proceed if it is coupled with another process which is so highly exergonic that the total value of ΔF for the combined reaction is negative. Such coupled reactions are of profound importance in biochemistry, and we now proceed to consider one of them in some detail.

THE STANDARD FREE ENERGY OF HYDROLYSIS OF ADENOSINE TRIPHOSPHATE

The supply of free energy in many coupled biochemical reactions is provided by the hydrolysis of adenosine triphosphate (ATP) to adenosine diphosphate (ADP) and inorganic phosphate. In solutions alkaline to pH 7, the reaction can be written (denoting the adenine and ribose groups, present in ATP and ADP, by the symbol A—R—)



or, in abbreviated notation,



However, ATP^{4-} can bind a hydrogen ion (at one of the negatively charged oxygens on the terminal phosphate) to form ATP^{--} ; similarly ADP^{--} can form ADP^- , and HPO_4^{--} can form H_2PO_4^- . (Binding of hydrogen ions at the other negatively charged groups occurs only at very acid pH values, so that it may be neglected here.) These equilibria involving hydrogen ions must be considered in any complete formulation of the process, and we shall return to them later. Analytically, however, we determine the total ATP concentration, i.e., the sum of (ATP^{4-}) and (ATP^{--}), and denote it by (ΣATP). Similarly we define

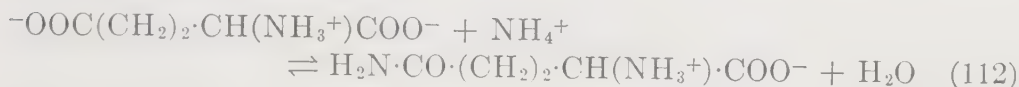
$$(\Sigma\text{ADP}) = (\text{ADP}^{--}) + (\text{ADP}^-)$$

and (ΣPO_4) = (HPO_4^{--}) + (H_2PO_4^-). In a process occurring at con-

stant pH , we may temporarily ignore the influence of hydrogen ion concentration, and write the process



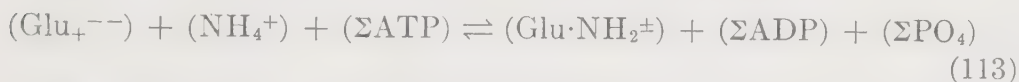
The equilibrium at any pH value lies very far to the right—so far that no detectable amount of remaining ATP can be observed. Thus the process denoted by (110) or (111) is highly exergonic. If we can use this process to drive another reaction which is endergonic, however, we may be able to achieve a true equilibrium. One system in which such an equilibrium has been attained involves the synthesis of glutamine from glutamate and ammonium ion:



or



The equilibrium in (112) lies far to the left—just how far, we shall consider shortly. If ADP, glutamine, and inorganic phosphate are dissolved together in the presence of an enzyme which has been extracted from green pea seeds, the hydrolysis of the glutamine is accompanied by a formation of ATP from ADP and inorganic phosphate according to the reaction which is equal to the sum of (111) and (112):



Experiments on this system were carried out by Levintow and Meister (1954), who showed that a true equilibrium was indeed established, the same final state being established whether one started with the reactants or with the products of (113). In one experiment, for instance, glutamine, ADP, and phosphate, each at a concentration of $10 \mu M$, were mixed in the presence of the enzyme at pH 7 and 37° . At equilibrium the concentrations of (NH_4^+) , glutamate, and ATP were all approximately $0.87 \mu M$, so that the concentration of each of the original reactants was $10 - 0.87 = 9.13 \mu M$. We then obtain the equilibrium constant, K_E , for reaction (113) at $310^\circ K$:

$$K_E = \frac{(Glu \cdot NH_2^\pm)(\Sigma ADP)(\Sigma PO_4)}{(Glu_{+}^{--})(NH_4^+)(\Sigma ATP)} = \left(\frac{9.13}{0.87} \right)^3 \cong 1.2 \times 10^3 = K_G K_P \quad (114)$$

Here K_G is the equilibrium constant of reaction (112), and K_P is that of reaction (111). The value of K_G has been determined by Benzinger

and Hems (1956), who followed, by a very delicate calorimetric technique the thermal changes which occurred when glutamate and ammonium ion at fairly high concentrations (0.2 to 0.9 *M*) were allowed to react in the presence of various amounts of added glutamine. An enzyme, glutaminase, from *Clostridium welchii*, was added to the system to catalyze reaction (112). If no glutamine, or only very minute amounts, were added to the system, reaction (112) proceeded from left to right, as was shown by absorption of heat in the calorimeter. If an excess of glutamine (say 0.001 *M*) was added at the start, reaction (112) proceeded from right to left, and heat was evolved. If just the right amount of glutamine was added at the start, equilibrium was achieved, and heat was neither evolved nor absorbed. In one experiment, for instance, this balance was attained with $(\text{Glu}_+^{--}) = (\text{NH}_4^+) = 0.884 \text{ M}$, $(\text{Glu} \cdot \text{NH}_2^\pm) = 0.00082 \text{ M}$. In such a system as this, the activity coefficients of the ions cannot be taken as unity without considerable error, because of the electrostatic interactions discussed in Chapter 5 below (see, for instance, Chapter 5, equation 76 and the immediately following equations). The mean ionic activity coefficient, already defined in equation (87), for the glutamate and ammonium ions was calculated as $f_\pm = 0.648$ for the system at equilibrium described above. The activity coefficient of the glutamine was taken as unity, since it is present at very low concentration and is moreover a dipolar ion, not an ion.¹⁰ Thus we may write

$$K_G = \frac{(\text{Glu} \cdot \text{NH}_2^\pm)}{(\text{Glu}_+^{--})(\text{NH}_4^+)f_\pm^2} = \frac{0.00082}{(0.884)^2(0.648)^2} = 0.00250 \quad (115)$$

$$\Delta F_G^\circ = -RT \ln K_G = 3500 \text{ cal mole}^{-1} \quad (298.1^\circ \text{ K})$$

The heat absorbed in the synthesis of glutamine was found to be $\Delta H_G = 5300 \text{ cal mole}^{-1}$. If we assume this value to be independent of temperature and take ΔH_G as equal to ΔH_G° , then (97) gives ΔF_G° at 310° K as 3430, or

$$K_P = K_E/K_G = (\Sigma \text{ADP})(\Sigma \text{PO}_4)/(\Sigma \text{ATP}) = 1200/0.00315 = 3.8 \times 10^5$$

$$\Delta F_P^\circ = -RT \ln K_P = -1419 \log K_P = -7900 \text{ cal mole}^{-1} \quad (116)$$

This is the quantity often referred to as the standard free energy of hydrolysis of adenosine triphosphate at pH 7. The value is probably the most reliable at present;¹¹ it is considerably lower than some of the values

¹⁰ If the standard state for glutamine is so defined that its activity coefficient is unity at infinite dilution in water without added salts, then its activity coefficient in 0.884 *M* ammonium glutamate may well be significantly less than unity (see the discussion of ion-dipole interactions in Chapter 5). The error involved here, however, is probably not large.

¹¹ Benzinger and Hems (1956) give the value as -7730 ; the difference between this and the value calculated in (116) is certainly within the limits of present experimental error.

estimated from earlier and more uncertain data, which were as high as $-10,000$ to $-12,000$ cal mole $^{-1}$.

The problem cannot be left at this point, however, for the value of K_E in equation (114) is a function of pH . Levintow and Meister (1954), for instance, found $K_E = 400$ at pH 6.0, and 3000 at pH 7.9. This variation is due to the fact that ATP, ADP, and inorganic phosphate all have pK values not far from 7, and equilibria involving proton uptake and release must be considered in order to describe the system adequately. On the other hand, K_G is virtually independent of pH , near pH 7, since the pK values of glutamic acid, glutamine, and ammonium ion all are either much greater or much less than 7 (see Chapter 8).

Alberty *et al.* (1951) have determined the following acid dissociation constants for the components of the equilibrium described by K_P in 0.15 M sodium chloride solution at 310°:

$$\begin{aligned} K_\alpha &= (H^+)(ATP^{4-})/(ATP^{3-}) = 10^{-6.50} \\ K_\beta &= (H^+)(ADP^{3-})/(ADP^{2-}) = 10^{-6.27} \\ K_\gamma &= (H^+)(HPO_4^{2-})/(H_2PO_4^-) = 10^{-6.73} \end{aligned} \quad (117)$$

Consider now a reaction involving certain specific ionic forms of ATP, ADP, and phosphate:



If we call its equilibrium constant K_L , then K_P in equation (116) may be expressed as the product of K_L , which is independent of pH , and a factor, Q , which is a function of (H^+) , K_α , K_β , and K_γ :

$$K_P = K_L Q = \frac{(ADP^{2-})(HPO_4^{2-})}{(ATP^{4-})} Q \quad (118)$$

$$Q = \frac{[1 + K_\beta/(H^+)] [1 + (H^+)/K_\gamma]}{[1 + (H^+)/K_\alpha]} \quad (119)$$

From (117) and (119) we obtain the values for Q given in Table V. From the value of K_P at pH 7 in equation (116), and the corresponding value of Q , we obtain the value of $K_L = 5.1 \times 10^3$ from (118). Hence $\Delta F_L^\circ = -6700$ cal mole $^{-1}$. Values for the standard free energy of splitting of ATP, ΔF_P° , as defined in (116), are listed in Table V. They are readily calculated from (118) and the listed values of Q .

At pH values above 8, (H^+) becomes so small that Q becomes practically equal to $K_\beta/(H^+)$, and ΔF_P° may be written

$$\begin{aligned} \Delta F_P^\circ &= -6700 - 1419[\log K_\beta - \log (H^+)] \\ &= -6700 - 1419(pH - 6.27) \\ &= 2200 - 1419 pH \quad (\text{for } pH > 8) \end{aligned} \quad (120)$$

From the data of Table V we may also readily calculate the variation of K_E (equation 114) with pH , since $K_E = K_P K_G$, and K_G is independent of pH between 6 and 8. At pH 6.0 we calculate $K_E = 380$, in excellent agreement with Levintow and Meister's experimental value of 400. At pH 7.9 we calculate $K_E = 7.2 \times 10^3$, much larger than their experimental value of 3×10^3 . This discrepancy is not surprising, however, since K_E at this point is so large that an error of a few per cent in the analyses could account for the entire difference. Considering the difficulties of the measurements, experiment and theory appear to be in good accord.

We note that the values of K_P and K_L , and the corresponding standard free energy changes, have been formulated without any attempt to introduce activity coefficients for ATP, ADP, or the phosphate ions.

TABLE V

STANDARD FREE ENERGY CHANGE, ΔF_P° , OF HYDROLYSIS OF ATP AS A FUNCTION OF pH , AT $310^\circ K$ ($37^\circ C$) (SEE EQUATION 116)

pH	Q	ΔF_P° (cal/mole)
6.0	2.37	-7200
6.5	3.65	-7600
7.0	7.5	-7900
7.5	19.3	-8500
7.9	45.0	-9000
8.0	56.0	-9200

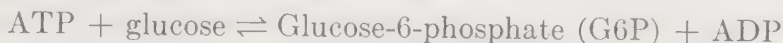
$$\Delta F_P^\circ = -RT \ln (K_L Q) = -1419(\log K_L + \log Q) = -6700 - 1419 \log Q$$

(See equations 118 and 119)

That is, we have taken the standard state for these ions as a solution of each at molar concentration, without distinguishing between concentration and activity. As we shall see in Chapter 5, the activity coefficient corrections for such highly charged ions are generally important, and the equilibrium constants K_P and K_L expressed in terms of concentration may be quite sensitive to variations in the ionic strength of the medium. Moreover ATP, ADP, and inorganic phosphate tend to form complexes with divalent, and even to some extent with monovalent, cations (Smith and Alberty, 1956a, 1956b). The functioning of the enzyme system used by Levintow and Meister required the presence of Mg^{++} or Mn^{++} ions, which must therefore to some extent have entered into complex formation with ATP and its breakdown products. Levintow and Meister, however, observed no change in K_E on varying the concentration of Mg^{++} in the reaction mixture, or on replacing Mg^{++} by Mn^{++} . More refined measure-

ments in the future, however, may show that these complicating factors must also be considered.

It is gratifying that a virtually identical value of ΔF_P° has recently been calculated by an entirely independent method. Robbins and Boyer (1957) have carefully determined the equilibrium constant for the phosphorylation of glucose by ATP, which proceeds in the presence of the enzyme hexokinase, according to the reaction



for which the equilibrium constant may be written

$$K_H = \frac{(\Sigma \text{ADP})(\Sigma \text{G6P})}{(\Sigma \text{ATP})(\text{glucose})} = 386 \text{ at } pH \text{ 6.0, } 30^\circ \quad (121)$$

The acid dissociation constant K_a for the phosphate in glucose-6-phosphate is $10^{-6.03}$; this is the only ionization which has to be taken account of for this compound between pH 6 and 8. Robbins and Boyer measured K_H at pH 6.0 and 30° , using an isotope dilution technique to determine the hexose concentration at equilibrium. From the ionization constants of reactants and products they calculated values of K_H at other pH values, by the same type of procedure we have already illustrated for the equilibria in the ATP-glutamine system. They thus calculated $K_H = 2460$ at pH 7.0, and 23,500 at pH 8.0. This gives, for the standard free energy change at pH 7.0,

$$\Delta F_H^\circ = -RT \ln K_H = -1394 \log K_H (pH \text{ 7, } 30^\circ) \\ = -4700 \text{ cal mole}^{-1} \quad (122)$$

Earlier, Meyerhof and Green (1948) had determined the equilibrium constant for the hydrolysis of glucose-6-phosphate to glucose and inorganic phosphate:

$$K_{G6P} = \frac{(\text{glucose})(\Sigma \text{PO}_4)}{(\Sigma \text{G6P})(\text{H}_2\text{O})} = 122 \text{ at } pH \text{ 8.5 and } 38^\circ \quad (123)$$

The activity of water has to be inserted explicitly into (123), since Meyerhof and Green worked with solutions which were very concentrated in glucose. In certain experiments, for instance, they started with 4.2 M glucose and 0.75 M phosphate, so that the concentration of water was only 27.5 moles per liter, instead of 55.5 as in pure water. The activity of water in such a system was assumed to be equal to $27.5/55.5 = 0.495$, taking the activity as unity for pure water, according to the usual convention for its standard state.¹² The standard state for the other compo-

¹² This assumes that the activity of water in these solutions is proportional to its volume concentration. No vapor pressure measurements on water in these systems are available to check this assumption, but it appears a reasonable approximation in the absence of vapor pressure data.

nents is taken as unit molarity in aqueous solution. Meyerhof and Green worked at pH 8.5. To calculate K_{G6P} at other pH values, Robbins and Boyer used the equation

$$K_{pH} = K_{pH\ 8.5} \times [1 + (H^+)/K_\gamma][1 + (H^+)/K_\delta]^{-1} \quad (124)$$

where $K_\delta = 10^{-6.03}$, as we have seen above. They employed the value of $10^{-6.83}$ for K_γ , instead of $10^{-6.73}$ which we have given in equation (117). They also corrected for the temperature effect on K_{G6P} , to convert from the value at 38° in Meyerhof and Green's experiments to 30° . Finally they obtained for the standard free energy change corresponding to K_{G6P} at 30° and pH 7.0

$$\begin{aligned} \Delta F_{G6P}^\circ &= -RT \ln K_{G6P} \quad (pH\ 7, 30^\circ) \\ &= -3100 \text{ cal mole}^{-1} \end{aligned} \quad (125)$$

The product of K_H (equation 121) and K_{G6P} (equation 123) is immediately seen to be equal to K_P , as defined by (111), (114), and (118). Hence for the standard free energies, from (122) and (125),

$$\begin{aligned} \Delta F_P^\circ &= \Delta F_H^\circ + \Delta F_{G6P}^\circ = -4700 - 3100 \\ &= -7800 \text{ cal mole}^{-1} \quad (pH\ 7, 30^\circ) \end{aligned} \quad (126)$$

It will be seen that this is in excellent agreement with the value calculated in Table IV by an entirely independent method.

We may note that Robbins and Boyer carefully considered the effect of added Mg^{++} ions on the equilibria, due to the binding of Mg^{++} to ATP and ADP in their different ionic forms. The values given above, however, are for systems in which (Mg^{++}) was very low, just enough to make the enzyme catalytically active, and its presence did not appreciably affect the equilibria.

It will be seen, from the examples considered here, that the calculation of free energy changes in biochemical systems is likely to be more complicated than in most of the simpler systems which have been studied by physical chemists in the past. In every case we have discussed, for example, it has been necessary to take account of acid-base equilibria in order to formulate adequately the free energy changes in the reactions. It was also necessary in all these cases to add an enzyme to the system in order to attain equilibrium. Indeed this has been one of the great contributions of biochemistry to thermodynamic studies; great numbers of equilibrium constants are readily determined, in systems containing enzymes, which could not be directly measured otherwise. The use of thermodynamic methods, on the other hand, has fundamentally clarified many biochemical problems, by showing which reactions could proceed spontaneously and which could proceed, if at all, only by the supply of free energy from

another source. The fact that a reaction can proceed spontaneously does not prove that it does so; its velocity may be zero, to all intents and purposes, unless an enzyme or some other catalyst is provided to overcome the resistances which otherwise prevent it from occurring. These problems of reaction velocity lie outside the scope of thermodynamics as we have presented it here;¹³ but in spite of this limitation the determination of free energy changes in biochemical systems is fundamental for the progress of biochemistry.

Systems Involving Other Variables

ELASTIC BODIES

Thus far we have limited ourselves to the discussion of systems for which pressure and temperature are the only physical variables that need be taken into account. It is often the case, however, that other physical variables are of significance also. A simple example is provided by an elastic body such as a piece of rubber, which may be stretched or compressed. Here the state of the body is determined not only by pressure and temperature but by tension also. Let us see how this case may be dealt with. We shall suppose that the body is a pure substance so that no composition variables come into question. If we introduce f to denote tension (a negative value of f will, of course, mean compression), then the work done *by* the body as a result of a reversible change of length, dl , is given by $-f dl$. If we go back to fundamentals we see that this expression will appear as an additional work term in (22), dW' being retained to refer to any work over and above pressure-volume *and* tension-length work done by the system, and that it will carry through in all the various developments of that equation. As a result we have as a general expression for the variation of E expressed as a function of S , V , and l , in place of (37),

$$dE = T dS - P dV + f dl \quad (127)$$

Similarly, for F expressed as a function of P , T , and l , we have, in place of (29.1),

$$dF = -S dT + V dP + f dl \quad (128)$$

¹³ The application of thermodynamic concepts to some irreversible processes has actually been achieved, with important results, beginning with the pioneer work of L. Onsager, shortly after 1930. Moreover, modern theories of reaction kinetics have been developed with the aid of ideas derived from thermodynamics—in particular, the concept of the formation of a critical complex, which is (often, but not always) in thermodynamic equilibrium with the reactants, and which decomposes to form the products of the reaction. These developments, however, lie beyond the scope of our discussion here.

These two equations show that

$$f = \left(\frac{\partial E}{\partial l} \right)_{S,V} = \left(\frac{\partial F}{\partial l} \right)_{P,T}$$

In addition to these, by taking account of the fact that the various second-order cross derivatives are equal to one another, we can derive a number of other significant relations from (127) and (128). For example, from (128) we obtain

$$- \left(\frac{\partial S}{\partial l} \right)_{P,T} = \left(\frac{\partial f}{\partial T} \right)_{P,l} \quad (129)$$

$$\left(\frac{\partial f}{\partial P} \right)_{T,l} = \left(\frac{\partial V}{\partial l} \right)_{T,P} \quad (130)$$

In order to obtain relations involving derivatives of l , we may conveniently rewrite (128) in an alternative form by subtracting $d(fl)$ from both sides. This gives

$$d(F - fl) = -S dT + V dP - l df \quad (131)$$

The quantity in parentheses on the left, which is a kind of extended free energy that takes account of the additional variables f and l , is, of course, an exact differential. Consequently we obtain at once the relations

$$\left(\frac{\partial l}{\partial P} \right)_{T,f} = - \left(\frac{\partial V}{\partial f} \right)_{T,P} \quad (132)$$

$$\left(\frac{\partial l}{\partial T} \right)_{P,f} = \left(\frac{\partial S}{\partial f} \right)_{T,P} \quad (133)$$

Many other such relations could be obtained by similar procedures, involving other variables.

In an earlier section we called attention to a particular feature of any substance which obeys the gas law, or indeed any substance for which pressure at constant volume is proportional to absolute temperature. This is that the energy depends only on temperature so that when the substance expands isothermally all the work done is accounted for by an entropy change and a corresponding absorption of heat. Owing to the exact correspondence of $-f$ and p and of V and l as variables expressing work, it is clear that a similar principle must hold in the case of an elastic body. If at any length (and pressure) tension is proportional to absolute temperature, then the energy of the body depends only on temperature, and any work done in an isothermal change of length is accounted for by an entropy change. The only exception will be if the volume of the body at constant temperature is dependent on extension, i.e., if stretching leads

to a change of volume. This at most will be a very minor effect. The proof of the proposition just stated is as follows. From (127)

$$\left(\frac{\partial E}{\partial l}\right)_{T,P} = T \left(\frac{\partial S}{\partial l}\right)_{T,P} - P \left(\frac{\partial V}{\partial l}\right)_{T,P} + f \quad (134)$$

If, now, we suppose

$$f = \phi(l, P)T \quad (135)$$

then by (129)

$$\left(\frac{\partial S}{\partial l}\right)_{T,P} = -\phi(l, P)$$

and, therefore, the first and third terms on the right cancel. Consequently

$$\left(\frac{\partial E}{\partial l}\right)_{T,P} = -P \left(\frac{\partial V}{\partial l}\right)_{T,P} \quad (136)$$

or alternatively, by (130),

$$\left(\frac{\partial E}{\partial l}\right)_{T,P} = -P \left(\frac{\partial f}{\partial P}\right)_{T,l} \quad (137)$$

We see, therefore, that if f is independent of P , and consequently V is independent of l , the energy depends only on the temperature. Many elastic polymers exert a tension which at any length is proportional to the absolute temperature, and, as a first approximation at least, independent of pressure. For such substances the energy must be essentially independent of length, and any work done as a result of a change of length must be accounted for by an entropy effect. This is a matter which will be discussed later in connection with the subject of diffusion. It is also of great importance in analyzing the properties of contractile systems such as muscle.

SYSTEMS IN GRAVITATIONAL OR CENTRIFUGAL FIELDS

Let us now turn to the case of a system in a gravitational or centrifugal field such that at any point in the field the system is acted on by a force proportional to its mass and to some function of its position. For simplicity, we shall suppose that the force acts always in one direction, which we identify with the x -axis of a set of rectangular coordinates. Then if the system moves through a distance dx , the work done by it is given by $+MG(x)dx$, where M is the total mass of the system and $G(x)$ is a function of x which gives the strength of the field. This expression for the work done is equivalent of $-f dl$ in the case of the elastic body, and, like the latter, will appear as an additional term in the fundamental

equation (22) and be represented in the various developments of that equation. As a result, equation (44) becomes

$$dF = V dP - S dT - MG(x)dx + \sum \mu_i dn_i \quad (138)$$

Equation (138) is a source of much information regarding a variety of partial derivatives. To begin with, it shows that

$$\left(\frac{\partial F}{\partial x}\right)_{P,T,n_i} = -MG(x) \quad (139)$$

and $(\partial F/\partial P)_{T,x,n_i}$ and $(\partial F/\partial T)_{P,x,n_i}$ remain equal to V and $-S$, as before. Next, if we put all variations, dn_i , equal to zero, it leads to a set of cross relations identical with those derived for the elastic body provided only we replace l by x and f by $-MG(x)$. The one corresponding to (132) shows that, if $G(x)$ were dependent on temperature, then the entropy of the system would vary with x . Finally, equation (138) gives us expressions for the variation of the chemical potentials with P , T , and x . Of these, the variations with P and T remain what they were in the simpler case where only P and T were considered. The variation with x , which is a new derivative, is given by¹⁴

$$\left(\frac{\partial \mu_i}{\partial x}\right)_{T,P,n} = -\left(\frac{\partial MG(x)}{\partial n_i}\right)_{T,P,x,n_j} = -G(x) \left(\frac{\partial M}{\partial n_i}\right)_{n_j} \quad (140)$$

It is evident in connection with (138) that, for any given values of P , T , and x , F is a first-order homogeneous function of the composition variables, n_i —that is to say, F is an extensive property of the system. Consequently, just as in the simpler case where only P and T were involved,

$$F = \sum n_i \mu_i \quad (141)$$

When this is differentiated and combined with (138), we obtain as an extended form of the Gibbs-Duhem relation (47),

$$\sum n_i d\mu_i = +V dP - S dT - MG(x)dx \quad (142)$$

This, at constant P , T , and x , passes over into

$$\sum n_i d\mu_i = 0 = \sum N_i d\mu_i \quad (143)$$

which is identical with (48). It should be realized that this extension of the relations derived for the simple case of a system dependent only on P and T would not always be permissible. It is only because of the

¹⁴ The subscript n_i on the left-hand side of (140) means that all the composition variables, n_i , are held constant. The subscript n_j on the right means that all the composition variables *except* n_i are held constant.

particular nature of the physical situation that the transition from (138) to (141) is valid; namely because F is a first order homogeneous function of the composition variables for constant values of P , T , and x . In the case of an elastic body where the work term $[MG(x)]$ is replaced by $-f dl$ such a simple transition is not possible.

Up to now it has been tacitly assumed that the system is small enough in relation to the scale of things so that we can specify its position by a single value of x and treat P and $G(x)$ as constant throughout the system. Let us now suppose that the system consists of an extended fluid. In this case, there will in general be a change of $G(x)$ within the compass of the system, and even in the special case where $G(x) = \text{constant}$ there will be a change of P with x due to hydrostatic effects. In a multicomponent system we must expect a change of composition with x . An analysis of the conditions of equilibrium in such a fluid leads us straight to the hydrostatic law and to the hypsometric equation for the distribution of components in a gravitational or centrifugal field, which are both matters of considerable importance.

Owing to the nonuniformity of the fluid in the x direction, it cannot be treated as a one-phase system. Instead we must think of it as composed of an infinity of phases, each phase consisting of an indefinitely thin slab perpendicular to the x -axis. Within each slab, pressure and composition will be constant and will correspond to a definite value of x and $G(x)$. Just as in the simpler case of a system composed of several phases each at constant temperature and pressure, the condition of equilibrium is that the chemical potential of each component be the same in each phase. In the case of the fluid, consisting of an infinite number of continuously varying slabs, this means that the total variation of the chemical potential of each component in the x direction must be zero. This variation is the resultant of the effects of composition, pressure, and $G(x)$. The condition of equilibrium is, therefore, represented by i equations, one for each component, each of the form

$$(d\mu_i)_{P,T,x} + \left(\frac{\partial\mu_i}{\partial P}\right)_{T,x,n} dp + \left(\frac{\partial\mu_i}{\partial x}\right)_{T,P,n} dx = 0 \quad (144)$$

The leading term represents the change in μ_i at constant P , T , and x due to composition changes. The differential coefficient in the second term is equal to \bar{V}_i , the partial molal volume of component i . That in the third term is, by (140), equal to

$$-G(x) \left(\frac{\partial M}{\partial n_i}\right)_{P,T,x} \quad (145)$$

The expression $(\partial M/\partial n_i)$ gives the change in total mass of the system per mole of the i th component added to the system. This is nothing but the molecular weight of that component, M_i . Consequently we may rewrite equation (144) as

$$(d\mu_i)_{P,T,x} + \bar{V}_i dP - M_i G(x) dx = 0 \quad (146)$$

There are i of these equations, one for each component. In order to obtain the hydrostatic relation we simply multiply each of these by N_i , the mole fraction of the corresponding component at any value of x , and add them.

Now by (143)

$$\Sigma N_i (d\mu_i)_{P,T,x} = 0$$

Also by (41)

$$\Sigma N_i \bar{V}_i = V$$

the volume of the phase per mole; and $\Sigma N_i M_i$ gives the mean molecular weight. Consequently $\Sigma N_i M_i / \Sigma N_i \bar{V}_i$ gives the density of the fluid for any value of x . We have, therefore, as a result of the operation,

$$dP = +\rho G(x) dx \quad (147)$$

This is the basic hydrostatic equation. It tells us that the increase of pressure with increasing x is equal to the density of the fluid multiplied by the quantity $G(x)$ which expresses the strength of the gravitational or centrifugal field. If we are concerned with the effect of gravity at the earth's surface, $G(x)$ is a constant equal to $-g$, the force of gravity, and

$$\frac{dP}{dx} = -g\rho \quad (148)$$

In applying (147), it is necessary to be clear as to the question of sign. By writing the work term as $MG(x)dx$, we have introduced the convention that the force acting on the fluid is positive when it leads to an increase of x . The hydrostatic equation is generally thought of as a mechanical principle and derived from mechanical considerations. It is interesting to see that it may also be regarded as a thermodynamic principle.

In order to obtain the hypsometric law we make use of the fundamental hydrostatic equation just deduced to express dP in terms of dx in the equations (146) expressing the constancy of the chemical potential of each component. The result is

$$(d\mu_i)_{P,T,x} - G(x)(M_i - \bar{V}_i \rho) dx = 0 \quad (149)$$

This cannot, of course, be integrated unless we have additional knowledge of the nature of the variation of μ_i with composition and the variation of \bar{V}_i , ρ , and $G(x)$ with x .

Perhaps the simplest application of (149) is to a liquid solution. In this case neither \bar{V}_i nor ρ will be very sensitive to pressure, and one may assume them to be constant. To simplify matters, we shall suppose that over the range of concentration the activity of the solute is proportional to its concentration (P , T , and x being fixed.) This makes it possible to write

$$(d\mu_i)_{P,T,x} = RT d \ln a_i = RT d \ln c_i$$

Equation (149) then becomes

$$RT d \ln c_i = (M_i - \bar{V}_i \rho) G(x) dx$$

which may be integrated to give the concentrations c_2 and c_1 at heights x_2 and x_1 :

$$\frac{c_2}{c_1} = \exp \left[\frac{(M_i - \bar{V}_i \rho)}{RT} \int_{x_1}^{x_2} G(x) dx \right] \quad (150)$$

In the case of a centrifugal field, we may identify x with the distance from the center of rotation. Then $G(x)$ becomes $\omega^2 x$, where ω denotes angular velocity. Consequently

$$\int_{x_1}^{x_2} G(x) dx = \frac{\omega^2}{2} (x_2^2 - x_1^2) \quad (150.1)$$

when this is substituted in (150), it shows how concentration increases with increasing x . This is the equation for sedimentation equilibrium in the ultracentrifuge. It is most commonly written in terms of the partial specific volume, \bar{v}_i , of component i , which is of course equal to \bar{V}_i/M_i . Since the unknown quantity to be determined in an ultracentrifuge run is generally the molecular weight of component i , we rearrange (150), after taking the logarithms of both sides, to express M_i in terms of the other quantities:

$$M_i = \frac{2RT \ln (c_2/c_1)}{\omega^2 (1 - \bar{v}_i \rho) (x_2^2 - x_1^2)} \quad (150.2)$$

Another significant application of the ideas we have been discussing is to the distribution of gases in the earth's atmosphere. For the purposes of this discussion, these gases may be assumed to be perfect without significant error. In the case of a mixture of perfect gases, we may express the total pressure as equal to the sum of the partial pressures, and the gas law becomes

$$PV = \Sigma n_i RT \quad (151)$$

Consequently the partial molal volume of the i th component is

$$\bar{V}_i \equiv \left(\frac{\partial V}{\partial n_i} \right)_{P,T} = \frac{RT}{P}$$

At the same time the partial pressure of the component is given by $N_i P$, and the chemical potential is $RT \ln N_i P$. Therefore

$$(d\mu_i)_{T,P,x} = RT d \ln N_i$$

In this case, however, instead of trying to integrate (149), it is easier to go back to (146), which now becomes

$$RT d \ln N_i + \frac{RT}{P} dP - M_i G(x) dx = 0 \quad (152)$$

or

$$RT d \ln N_i + RT d \ln P - M_i G(x) dx = 0$$

If we set $G(x) = -g$, equation (152) may be integrated to give

$$RT \ln N_i P + M_i g x + \text{Const.} = 0$$

Since $N_i P$ is, of course, simply the partial pressure, p_i , of the i th component, this may be written as

$$RT \ln p_i = -M_i g x + \text{Const.}$$

or

$$p_i = p_{i_0} \exp \left[- \frac{M_i g (x - x_0)}{RT} \right] \quad (152.1)$$

where p_{i_0} is the partial pressure of component i at $x = x_0$, say at sea level. This is the most familiar form of the hypsometric law.

Statistical Interpretation of Entropy

In a general way, it has been long recognized that highly ordered systems have low entropy and that an increase of disorder corresponds in some way to an increase of entropy. Thus the melting of a crystal is accompanied by an increase of entropy, corresponding to the transition from the highly ordered structure of the crystal to the more random arrangement of the molecules in the liquid. Numerically the entropy increase is given, per mole of substance, by the heat absorbed during melting, divided by the absolute temperature. The increase in the randomness of molecular distributions, which occurs when a liquid is vaporized, is likewise reflected in the positive entropy change on vaporization. Another example is the isothermal mixing of two gases—for instance, a tankful of oxygen with a tankful of nitrogen. If the gases are dilute, the process occurs without appreciable absorption or evolution of heat. Before mixing occurred, we could specify that any molecule of oxygen in the system was in tank A, not in tank B; after mixing, this particular piece of information is lost to us; the constituents of the system have become more randomly distributed than before, and the entropy has increased.

The maximum degree of order conceivable is probably that of a perfect crystal at the absolute zero of temperature. In such a system, once we know the position of a single atom and the structure and orientation of the crystal, it is possible to specify the position of all the other atoms, assuming them to be at rest at 0° K. As such a crystal absorbs heat, the component atoms and molecules of the crystal are set into vibration. Such vibrations occur, in an increasingly large number of the atoms or molecules composing the crystal, as the temperature rises. Thus the system departs more and more from its original uniquely defined configuration at the absolute zero; the number of possible configurations compatible with the total energy of the crystal increases with increasing temperature, and this fact is intimately related to the increase of the entropy which can be calculated from measured values of the heat capacity of the crystal as a function of temperature (see equation 33). When it melts there is a further large increase in the number of possible configurations. When the liquid vaporizes there is another and generally much larger entropy increase corresponding to the greatly increased randomness in the distribution of molecules in a gas as compared to that in a liquid.

The problem is to find a suitable way of expressing this degree of randomness in quantitative terms. Our knowledge of the system as a whole imposes certain specifications which any picture of the molecular configurations must fulfill. We can determine experimentally the energy (E) of the system—measured relative to some suitably chosen reference state for which the energy is taken as zero. We can also determine its volume (V) and the total number of molecules (N_i) of each of the different kinds of chemical components of which it is composed. Specification of energy, volume, and composition determines the temperature of the system, or, conversely, the energy is determined if pressure, temperature, and composition are specified. Subject to these conditions, however, it would still be possible to construct an enormous variety of possible configurations of the individual molecules, each of which would give the specified energy, volume, and composition. We may call each such configuration a microscopic configuration, as distinguished from the single macroscopic configuration of the system as a whole, which is specified by E , V , and the N_i 's. The total number of microscopic configurations compatible with a given macroscopic configuration is obviously a very large number. This number, however, though large, is finite, for the study of quantum mechanics has shown that molecular vibrations and rotations are restricted to certain discrete energy levels. Even in the case of translational velocities and momenta, for which physicists of earlier generations would certainly have supposed that a continuum of possible values

would be attainable, the laws of quantum mechanics restrict the possible values to a series of discrete translational energy levels. Such restrictions do not apply to particles which are free in open space, but particles confined within a container are definitely so restricted. We can describe a possible microscopic state of the system by specifying the number of molecules of each kind in the various levels, subject to the general requirement that the total energy and the total number of molecules of each kind in the system must add up to the correct values, and that they must all be confined in the specified volume.

Thus, if we specify a given macroscopic state of the system by the total values of E and V , and of the N_i 's, it is possible to calculate the total number, Ω , of microscopic configurations which are compatible with the given macroscopic state. The number Ω is intimately related to the entropy. Qualitatively, it increases as the entropy increases. It is inherently a positive number; hence its logarithm is positive or (in the limit when $\Omega = 1$) zero. Often Ω is referred to as a measure of the probability of finding a system in a given macroscopic state. This use of the probability concept implies that any one microscopic distribution is just as probable as any other; the high probability of a given equilibrium state of a macroscopic system corresponds to the fact that many distributions (microscopic configurations) are compatible with that state.

Before considering how the number Ω may actually be evaluated for specific systems, we may see readily the general form of the relation which must exist between Ω and the entropy of the system. Entropy is an extensive property; if we have one bottle containing a system A with entropy S_A under specified conditions, and another bottle containing a system B with entropy S_B , then the entropy of the two systems is $S_A + S_B$. The system A may be in any one of Ω_A microscopic states, the system B in any one of Ω_B states. Any one of the Ω_A states of the former may coexist with any one of the Ω_B of the latter. Hence the number of possible states of the larger system which includes both A and B is $\Omega_A \Omega_B$. Thus $\Omega_{A+B} = \Omega_A \Omega_B$; or $\ln \Omega_{A+B} = \ln \Omega_A + \ln \Omega_B$. Thus, it is the logarithm of Ω , not itself, which is an extensive property of any system; and the expected relation between S and Ω is of the form

$$S = k \ln \Omega \quad (153)$$

The factor k in equation (153) turns out in fact to be Boltzmann's constant k , the gas constant per molecule; and $k = R/N = 1.380 \times 10^{-16}$ ergs per degree.*

*Analogous considerations lead to the concept of the entropy of language in modern information theory. The amount of information in a book of two volumes is twice as great as that in a book of one volume, but the number of different meaningful

From a knowledge of the possible energy levels—translational, rotational, vibrational, and electronic—which the molecules of a system can assume, in accordance with the principles of quantum mechanics, one may calculate the statistical distribution of the molecules among the various levels at equilibrium. The translational energy levels are determined by the masses of the molecules and the size of the container; the other energy levels can be determined from spectroscopic measurements. Exact calculations can generally be made at present only for systems of isolated molecules, i.e., for dilute gases. For most such systems—indeed for practically all which concern us in this book—the distribution of molecules among the different energy levels is given by the Boltzmann distribution law:

$$N_i = N_0 e^{-(\epsilon_i - \epsilon_0)/kT} \quad (154)$$

Here N_0 is the number of molecules in the ground state of lowest energy ϵ_0 , and N_i is the number of molecules of energy ϵ_i , above the ground level.¹⁵

Since the possible energy levels are quantized, and since the energy of the total system is finite, and specified, the number of ways in which it is possible to distribute the molecules among all the different energy levels is also finite. This number is Ω , and its natural logarithm, multiplied by k , gives the entropy according to (153).

Moreover it is possible, once the various molecular energy levels are known, to obtain values for the other thermodynamic functions of the system also. The distribution law (154) or (154.1), by specifying the relative number of molecules in each of the possible energy levels, permits a summation over all the possible states; we multiply the number of molecules in each level by the energy per molecule in that level, and add all these terms together to obtain the energy of the entire system. By further calculations we can obtain the heat capacity, heat content, free energy, and other thermodynamic functions of the system, once we know the possible energy levels derived from spectroscopic measurements. The discussion of such measurements lies far outside the scope of this book;

arrangements of words possible in the two volumes is the square of the number possible in the one volume. Information is therefore related logarithmically to the number of possible arrangements of words, just as the physicist's entropy is related logarithmically to the number of possible physical configurations. Different languages, particularly languages of different groups such as English and Arabic, differ characteristically when considered from this point of view. See for instance Norbert Wiener "Cybernetics."

¹⁵ If there are g_i different quantum states which all have the same, or nearly the same, energy, ϵ_i , then this level of energy is said to be g_i -fold degenerate, and the Boltzmann equation becomes:

$$N_i = N_0 g_i e^{-(\epsilon_i - \epsilon_0)/kT} \quad (154.1)$$

the reader is referred to treatises on statistical mechanics. (See, for example, Mayer and Mayer, 1940; Rushbrooke, 1949; Fowler and Guggenheim, 1939. A relatively brief and incomplete, but very useful, discussion is given by Glasstone, 1947, Chapters VI and IX; a treatment intermediate between this and that of the more extensive treatises is given by Denbigh, 1955.) We merely note here that many of the most accurate data on the thermodynamic functions for simple molecules are derived from such measurements.

The calculation of entropy from (153) is in general outside the scope of our discussion, but we may illustrate the use of this equation by a few simple examples.

ENTROPY OF A PERFECT CRYSTAL AT 0° K

As our previous discussion (p. 225) has indicated, a perfect crystal should be in a uniquely defined configuration, that is in a single quantum state at the absolute zero of temperature. In this case Ω should assume its lowest possible value of 1, and the entropy should be zero. This postulate is the essence of what is commonly called the Third Law of Thermodynamics, which has been stated in various, more or less equivalent ways. It was first formulated in a rather incomplete fashion, early in the twentieth century, as a result of the work of T. W. Richards, W. Nernst, M. Planck, and others, but it could not be given an adequate theoretical basis until the theory of quantum phenomena was systematically developed.

The thoughtful reader will note that we have not defined a perfect crystal, and will wonder how he is to know one when he sees it. This is indeed a very troublesome question, and we are generally forced to measure the entropy of the crystal in order to get the answer. The entropy of the substance in the vapor state, say at 298.1° K, may be evaluated, given the knowledge of the molecular energy levels from spectroscopic studies. The *increase* in entropy, from 0° K up to the same condition in the vapor state at 298.1° K, may be determined from heat capacity and other thermal measurements (see equation 162 below). If we get the same numerical value by both methods, then we are justified in calling the crystal perfect from the point of view of the Third Law. In some cases, however, the entropy as calculated from spectroscopic data proves to be higher than that calculated from thermal data (it is never lower). This indicates that there is some residual entropy, some randomness of arrangement, in this particular crystal even at 0° K; it is not a perfect crystal. An example of such an imperfect crystal is ice, which is discussed under section 3 below. We may note, however, that the vast majority of

organic and inorganic crystals may justifiably be regarded as perfect by this criterion.

ENTROPY CHANGE ON MIXING OF COMPONENTS IN AN IDEAL SOLUTION

The numerical use of the statistical conception of entropy may be illustrated by the calculation of the entropy change in the isothermal mixing of two components which form a perfect solution. We have already treated this problem by the methods of classical thermodynamics, and derived equation (64). Here we shall show that the statistical conception of entropy, using a simple model, leads to the same result. The identical formula is also found to apply to the mixing of two components in the gaseous state, if each of the components and the final mixture is a perfect gas. Likewise the formula applies to mixtures in a perfect crystal, such for instance as may occur if the crystal is made up of two different kinds of molecules which are identical except that one contains a different kind of isotope of one particular atom from the other.

As we have already seen (p. 184), a perfect solution is defined as one in which the activity of each component is proportional to its mole fraction and in which there is no heat or volume change on mixing. If a solution containing two components is to meet these requirements, the two kinds of molecules of which it is composed, which we may denote as type 1 and type 2, respectively, must be of nearly the same size and shape, so that we can remove a type 1 molecule and replace it by a type 2 molecule without altering the volume of the solution. Furthermore, the requirement that there shall be no heat of mixing implies that the interactions between a type 1 and a type 2 molecule are essentially the same as those between 1 and 1 or between 2 and 2. If this requirement is fulfilled, the heat of mixing should be zero within experimental error. The entropy of mixing then involves no thermal change but is due only to the increased randomness in the system resulting from the mixing. Let M_1 molecules of type 1 be mixed with M_2 molecules of type 2, and assume that the volume is just sufficient to accommodate the total number of molecules, $M_1 + M_2$. We may imagine this volume to be divided up into $M_1 + M_2$ small volume elements, as if it were a lattice, each volume element containing just enough space for one molecule.

If all the molecules were of one kind, and therefore indistinguishable, there would be just one way of fitting them into the $M_1 + M_2$ volume elements or holes. If there are two kinds of molecules present, however, we may start with the M_1 molecules of type 1 and insert them into the holes, one by one. If the molecules of type 1 were numbered and distinguishable, we could place the first molecule in any one of $M_1 + M_2$ holes,

the second in any one of $M_1 + M_2 - 1$ holes, and so forth. For the last of the M_1 molecules there would be $M_1 + M_2 - (M_1 - 1) = M_2 + 1$ possible positions to choose from. Thus the total number of possible arrangements, for distinguishable molecules, is

$$(M_1 + M_2)(M_1 + M_2 - 1)(M_1 + M_2 - 2) \cdots (M_2 + 2)(M_2 + 1) \\ = (M_1 + M_2)!/M_2!$$

The molecules, however, are not in fact distinguishable. Thus, if we have placed M_1 molecules in any *specified* set of M_1 volume elements, out of the total of $M_1 + M_2$ that are available, this corresponds to one microscopic configuration. In the counting given above, however, with imaginary distinguishable molecules, we have counted this one state as if it were $M_1!$ states; for we can arrange M_1 *distinguishable* molecules in a specified set of M_1 holes in $M_1!$ different ways. To get the true number, Ω , of new microscopic states that are made available to the system by the process of mixing, we must therefore divide the number given above by $M_1!$. Thus we obtain for Ω :

$$\Omega = \frac{(M_1 + M_2)!}{M_1!M_2!} \quad (155)$$

To complete the mixing process, for any specified microscopic configuration we must now fill in the M_2 unoccupied holes with molecules of type 2. Since the molecules of any one type are indistinguishable, there is only one possible way of doing this, and this process does not increase the value of Ω .

It should be noted that equation (155) is symmetrical, as it should be, with respect to the two kinds of molecules. If we had started by putting in M_2 molecules of type 2 into the $M_1 + M_2$ holes, and then inserted the type 1 molecules subsequently, we should have come out with the same answer.

Thus the extra entropy of the system which arises from mixing the type 1 and type 2 molecules is given by the equation

$$\Delta S = k \ln \Omega = k \ln \frac{(M_1 + M_2)!}{M_1!M_2!} \quad (156)$$

In practice we are dealing with numbers of molecules of the order of that contained in one gram mole—that is, of the order of 10^{23} . For any number, M , which is even moderately large, we may replace $M!$ by Sterling's approximation which in logarithmic form may be written with sufficient accuracy:

$$\ln M! = M \ln M - M \quad (157)$$

Inserting (157) to calculate the factorials in (156), and dividing the result by $M_1 + M_2$ to express the result as ΔS per molecule, we obtain:

$$\Delta S = -k \left[\frac{M_1}{M_1 + M_2} \ln \left(\frac{M_1}{M_1 + M_2} \right) + \frac{M_2}{M_1 + M_2} \ln \frac{M_2}{M_1 + M_2} \right] \quad (158)$$

We may note immediately that $M_1/(M_1 + M_2)$ is the mole fraction N_1 of component 1, and $M_2/(M_1 + M_2)$ is the mole fraction N_2 of component 2. If now we make the further choice that $M_1 + M_2$ is equal to Avogadro's number, N_0 , then we have the equation for entropy change on mixing, per mole of solution, since $R = kN_0$:

$$\Delta S = -R(N_1 \ln N_1 + N_2 \ln N_2) \quad (64)$$

This is, of course, identical with expression (64), derived by the methods of classical thermodynamics. The classical derivation has the advantage of leading to significant results with a minimum of assumptions. The derivation in terms of the statistical concept of entropy, in the form given here, involves the somewhat arbitrary concept of a lattice, within the framework of which we place the molecules. It provides us, however, with a molecular picture of the variety of distributions possible within the system, which opens the way to the interpretation of more complicated systems along similar lines. Other, more refined models may be introduced, which remove some of the crudities of our simplified lattice model. For instance, the modern theory of solutions of large flexible polymer molecules, which can coil and twist into a great variety of different forms, all of approximately equal energy, has been based largely on extensions of the type of analysis indicated here. For a discussion of such studies on polymers, with many references, the book of Flory (1953) may be consulted.

THE RESIDUAL ENTROPY OF ICE

In the discussion of water in Chapter 2, we noted that each oxygen atom in the ice crystal is surrounded by a tetrahedron of four other oxygens. Each oxygen atom has two hydrogens attached to it by covalent bonds and two pairs of unshared electrons, if we ignore the H_3O^+ and OH^- ions, which are extremely rare compared to the H_2O molecules. Each of the two H nuclei attached to any given oxygen atom points toward an unshared electron pair in one of the neighboring oxygens. The H nuclei, however, at ordinary temperature, are constantly leaping from one oxygen to another. The orientation of the two hydrogens which are attached to any given oxygen at a particular moment may thus correspond to any one of six possibilities; if we arbitrarily number the four surrounding oxygens, shown in Chapter 2, Fig. 3, as 1, 2, 3, and 4, the two H nuclei

on the central oxygen may point toward 1 and 2, 1 and 3, 1 and 4, 2 and 3, 2 and 4, or 3 and 4. In practice, however, not all six orientations are allowable; they are forbidden if the adjacent position on the neighboring oxygen is occupied by another proton, and are allowed only if there is an unshared electron pair in this orientation. The chance that a given oxygen, say oxygen 1, is thus available is $\frac{1}{2}$; the chance that any given two of the four oxygens are simultaneously available is $(\frac{1}{2})(\frac{1}{2}) = \frac{1}{4}$. Thus the *average* number of possible orientations for the two hydrogens around any given oxygen is $\frac{6}{4} = \frac{3}{2}$. This is the number of possible arrangements of H nuclei, per water molecule in the ice crystal, and the number per mole is $\Omega = (\frac{3}{2})^N$. Thus the contribution of this randomness of arrangement to the entropy of the system is

$$S_0 = k \ln (\frac{3}{2})^N = kN \ln (\frac{3}{2}) = R \ln (\frac{3}{2}) \\ = 0.81 \text{ cal deg}^{-1} \text{ mole}^{-1} \quad (159)$$

As the ice crystal is cooled down to lower and lower temperatures, the H nuclei cease to leap from oxygen to oxygen, and their distribution is "frozen" into one of the $(\frac{3}{2})^N$ possible arrangements, all of which have so nearly the same energy that no one is particularly favored over the others. Thus the ice crystal, when cooled to 0° K, is not a perfect crystal, since this residual randomness of arrangements of the H nuclei persists, and its entropy at 0° K may be taken from (159) as 0.81 cal deg⁻¹ mole⁻¹. This value has been checked experimentally, for the entropy as calculated from spectroscopic energy levels of the water molecule in the vapor state is higher than the entropy calculated from heat capacity measurements and heats of fusion and vaporization (equation 162 below). The discrepancy is equal, within the experimental error, to $R \ln (\frac{3}{2})$.

This explanation of the residual entropy of ice was originally proposed by Pauling (see Pauling, 1940, Chapter IX, where equation 159 is derived by reasoning somewhat different from, but equivalent to, that given here).

Standard Free Energies of Formation and Their Use in Determining Thermodynamic Equilibria

Many equilibrium constants, not readily accessible to direct experimental measurement, can be calculated, on the basis of the principles we have been discussing, from the heats of combustion of the compounds involved, and from entropies derived from heat capacity measurements and from heats of fusion and vaporization. The standard heat of formation (ΔH°_f) of a compound is the increase of heat content when one mole of the compound in its standard state, at the standard temperature and pressure—usually 25° and one atmosphere—is formed from its elements, each in its respective standard state. For carbon this standard state is

pure graphite; for hydrogen, oxygen, and nitrogen it is the pure gas at one atmosphere pressure. The details of calculating these heats of formation from heats of combustion will not be presented here; they are discussed in several of the references given at the end of this chapter (for instance, by Glasstone, by Klotz, and by Rossini).

Likewise we shall deal very briefly with the profoundly important problem of entropy calculations. If we measure the heat capacity of a crystalline solid, starting at a very low temperature, T_1 , close to 0°K , and gradually warming it up to the melting point at temperature T_m , the entropy increase of the solid during the process is, from (33),

$$S_2 - S_1 = \int_{T_1}^{T_m} (C_p/T) dT = \int_{T_1}^{T_m} C_p d \ln T \quad (160)$$

Here, of course, C_p is the heat capacity at constant pressure. It is found experimentally, and the modern quantum theory of solids, starting with the work of Einstein and Debye, predicts the finding theoretically, that C_p decreases rapidly and approaches zero at very low temperatures. Indeed it approaches zero so rapidly that C_p/T also approaches zero as T approaches zero.¹⁶ Consequently the integrand in (160) remains finite, even if the lower limit of integration is taken as 0°K —a limit which of course can never be attained experimentally, although it can be very closely approached. Hence the entire integral in (160) is a definite finite quantity. When the solid melts, at T_m , with an increase of heat content, ΔH_m , the entropy increase is, from the fundamental definition for an entropy change in a reversible process,

$$\Delta S_m = \Delta H_m/T_m \quad (161)$$

If we then warm up the liquid from the melting point to the boiling point, T_v , and measure the heat capacity as a function of temperature, we evaluate the entropy increase by an integral of exactly the same form as (160), except that the limits of integration are T_m and T_v . If, at T_v , the liquid is vaporized to give a gas at one atmosphere pressure, the entropy of vaporization is $\Delta S_v = \Delta H_v/T_v$. Thus the entropy of the vapor at T_v and one atmosphere is

$$S_{T_v} = S_0 + \int_0^{T_m} C_p d \ln T + \Delta H_m/T_m + \int_{T_m}^{T_v} C_p d \ln T + \Delta H_v/T_v \quad (162)$$

Here S_0 is the entropy of the solid at 0°K . Obviously the increase in entropy of the vapor, as the temperature rises above T_v , may be evaluated by more heat capacity measurements. If we wish to determine the en-

¹⁶ For more detailed discussions of this point, see for instance Rushbrooke (1949) and Glasstone (1947).

tropy increment from 0°K up to (say) 298.1°K , and if (for instance) the substance in question were a solid over this whole range, we should, of course, require only the first two terms on the right-hand side of (162).¹⁷ Thus, in order to fix the value of the entropy, the only thing that remains is to determine S_0 . Our previous discussion of the statistical conception of entropy indicates that for most pure substances we may take $S_0 = 0$. The chief exception of biochemical interest is water, for which we have already calculated $S_0 = R \ln (3/2)$ in (159). We may then tabulate the resulting entropy for the substance in its standard state at 298.1°K , denoted by $S_{298.1}^\circ$. Similar calculations have been made for the elements of which the substance is composed, and their entropies in their standard states are known. It is then a simple matter of subtraction to calculate the standard entropy of formation, $\Delta S^\circ f$, of the substance from its elements in their standard states. The standard free energy of formation, $\Delta F^\circ f$, is then, of course, immediately given by the equation

$$\Delta F^\circ f = \Delta H^\circ f - T \Delta S^\circ f \quad (163)$$

We may then obtain the standard free energy of a chemical reaction, ΔF° , by summing the values of $\Delta F^\circ f$ for all the products, and likewise for all the reactants, and subtracting the latter sum from the former.

The value of $\Delta F^\circ f$ is generally given for the formation of the pure substance from its elements, at 298.1°K and one atmosphere. For most purposes, however, it is more important to know the free energy of formation of the substance in aqueous solution at unit activity, using the convention that the activity approaches the molality at infinite dilution of the solute. This calculation is readily made from $\Delta F^\circ f$, if the activity of the solute is known as a function of concentration, and if its solubility has been determined. We may illustrate the calculation for the case of glycine, for which $\Delta F^\circ f$ at 298.1°K has been determined as -88.61 kcal/mole . The chemical potential of glycine in the saturated solution is, of course, the same as in the solid state, since the two are in equilibrium. There is therefore no change in free energy when solid glycine dissolves in the saturated solution. The change in chemical potential when one mole of glycine is transferred from saturated solution to a solution of unit activity in water is

$$\begin{aligned} \mu^\circ - \mu_{\text{sat}} &= RT \ln (1/a_{\text{sat}}) = -RT \ln a_{\text{sat}} \\ &= -2.303 RT (\log m_{\text{sat}} + \log f_{\text{sat}}) \end{aligned} \quad (164)$$

¹⁷ Some solids undergo transitions from one crystalline form to another, at certain sharply defined temperatures. These involve the absorption of heat (ΔH_t) if the transition occurs in the direction of the form which is stable at the higher temperature. The entropy change here is, of course, given by $\Delta S_t = \Delta H_t/T_t$, where T_t is the transition temperature.

For a saturated solution of glycine in water at 298.1° K, $m_{\text{sat}} = 3.3$, $\log m_{\text{sat}} = 0.52$, and we find from Table II that $\log f_{\text{sat}} = -0.137$. Hence $\log m_{\text{sat}} + \log f_{\text{sat}} = 0.38$, and

$$\begin{aligned}\Delta\tilde{F}^\circ - \Delta F^\circ f &= \mu^\circ - \mu_{\text{sat}} = -2.303RT(0.38) = -1365 \times 0.38 \\ &= -0.52 \text{ kcal/mole} \quad (164.1)\end{aligned}$$

Here we follow the notation of Borsook, denoting the standard free energy of formation of the solute at unit activity in solution by $\Delta\tilde{F}^\circ$. Thus $\Delta\tilde{F}^\circ$ for glycine at unit activity in water is $-88.61 - 0.52 = -89.13$ kcal/mole.

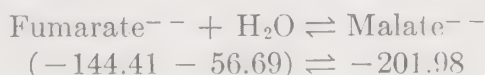
We note here that a glycine solution of unit activity is not exactly one molal, but slightly more concentrated. From Table II, we find $-\log f = 0.0684$ when $m = 1$. A solution of unit activity has a value of $\log m = 0.075$ very nearly, so m (unit activity) = 1.19.

In Table VI the standard free energies of formation of a number of substances of importance in biochemistry are listed. The free energies of a large number of reactions may be calculated from these data. It will be noted that $\Delta\tilde{F}^\circ$ values for a number of ions in aqueous solution at unit activity are also listed. It is, of course, impossible to add an ion to a solution, or remove it from solution, without also adding or removing other ions to balance the net charge on the ion in question. If we arbitrarily assign a value of $\Delta\tilde{F}^\circ$ for any one ion, however, we may then calculate values for all other ions from the experimental data. The convention universally adopted is to write $\Delta\tilde{F}^\circ = 0$ for H^+ ion at unit activity.

We may illustrate the use of Table VI by a few examples.

THE FUMARATE-MALATE EQUILIBRIUM

The reaction (see equation 98 and the accompanying text) is



The value of $\Delta\tilde{F}^\circ$ for each component of the reaction is written directly below its symbol. Subtracting the sum of the two terms on the left from the term on the right, we obtain $\Delta F^\circ = -0.88$ kcal/mole. This agrees with the value given in (99), as, of course, it should.

THE FIRST IONIZATION OF CARBONIC ACID

The reaction to be considered is

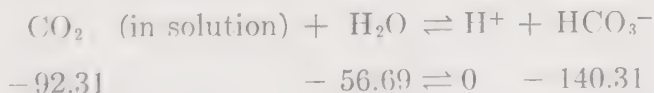


TABLE VI

STANDARD FREE ENERGIES OF FORMATION OF SOME IMPORTANT SUBSTANCES
IN THE PURE STATE ($\Delta F^\circ f$) AND IN AQUEOUS SOLUTION AT 1 M ACTIVITY
($\Delta \tilde{F}^\circ$)

(All values in kcal mole⁻¹ at 25° C, 298.1° K)

Substance	$-\Delta F^\circ f$	$-\Delta \tilde{F}^\circ$
Acetaldehyde (gas)	31.96	33.38
Acetic acid (liquid)	93.75	95.48
Acetate ⁻	—	88.99
Acetoacetate	—	118.00
Acetone (liquid)	37.18	38.52
<i>cis</i> -Aconitate ⁻⁻⁻	—	220.51
L-Alanine	88.40	88.75
D,L-Alanine	—	89.11
D,L-Alanylglycine	—	114.57
NH ₃ (gas)	3.98	6.37
NH ₄ ⁺ ion	—	19.00
L-Asparagine	183.50	125.86
	(monohydrate)	
L-Aspartic acid	174.76	172.31
L-Aspartate ⁺⁻	—	166.99
<i>n</i> -Butanol (liquid)	40.39	41.07
<i>n</i> -Butyric acid (liquid)	90.65	90.86
<i>n</i> -Butyrate ⁻	—	84.28
CO ₂ (gas)	94.26	92.31
HCO ₃ ⁻	—	140.31
Citrate ⁻⁻⁻	—	279.24
Isocitrate ⁻⁻⁻	—	277.65
Creatine	63.19	—
	(anhydrous)	—
	121.11	63.17
	(monohydrate)	
Creatinine	6.90	6.91
Cystine	163.55	159.00
Cysteine	82.08	81.21
Ethanol (liquid)	41.77	43.39
Formaldehyde (gas)	26.30	31.2
Fructose	—	218.78
Fumaric acid	156.49	154.67
Fumarate ⁻⁻	—	144.41
α -D-Galactose	220.00	220.73
α -D-Glucose	217.56	219.22
L-Glutamic Acid	173.81	171.76
L-Glutamate ⁺⁻	—	165.87
Glycerol (liquid)	114.02	116.76
Glycine	88.61	89.14
Glycogen (per glucose unit)	—	158.3
Glycolate ⁻	—	126.9

TABLE VI (Continued)

Substance	$-\Delta F^\circ_f$	$-\Delta F^\circ$
Glyoxylate ⁻	—	112.0
OH ⁻ ion	—	37.60
H ⁺ ion	—	0.00
H ₂ S (gas)	7.89	6.54
HS ⁻	—	-3.0
β -Hydroxybutyric acid	—	127.00
β -Hydroxybutyrate ⁻	—	121.00
Isocitrate ⁻⁻⁻	—	277.65
Isopropanol	43.26	44.44
α -Ketoglutarate	—	190.62
Lactate ⁻	—	123.76
α -Lactose	419.11	362.15
	(monohydrate)	
β -Lactose	374.56	375.26
	(anhydrous)	
L-Leucine	82.63	81.68
D,L-Leucine	—	81.76
D,L-Leucylglycine	—	110.90
Mannitol	225.20	225.29
Malate ⁻⁻	—	201.98
β -Maltose	413.48	357.80
	(monohydrate)	
Methanol (liquid)	39.73	41.88
NO ₃ ⁻	—	26.41
NO ₂ ⁻	—	8.25
Oxalacetate ⁻⁻	—	190.53
<i>n</i> -Propanol	41.21	42.02
Pyruvate ⁻	—	113.44
Succinic acid	178.68	178.39
Succinate ⁻⁻	—	164.97
Sucrose	369.20	370.90
L-Threonine	—	123.0
L-Tyrosine	96.10	92.55
Urea	47.12	48.72
Water (liquid)	56.69	—

Standard state for a gaseous component is the pure gas at 1 atm; for a liquid or solid, the pure substance standard state in aqueous solution is 1 *M* activity, setting activity of solute equal to molality as $m \rightarrow 0$.

Most of the values here listed are from a table compiled by K. Burton in Krebs and Kornberg (1957); see also Burton and Krebs (1953), Krebs (1954). Some of the values for amino acids and peptides are from Borsook (1953). Original sources are given in the references cited.

Again, subtracting the terms on the left from those on the right, we obtain $\Delta F^\circ = 8.69$ kcal/mole. Since ΔF° (cal/mole) = $2.303RT \log K_1$ for carbonic acid, this gives $\log K_1 = 6.36$. In Chapter 10 we shall consider this reaction further, and shall show also how its total free energy may be subdivided into the free energies of the component reactions: (1) $\text{H}_2\text{O} + \text{CO}_2 \rightleftharpoons \text{H}_2\text{CO}_3$ and (2) $\text{H}_2\text{CO}_3 \rightleftharpoons \text{H}^+ + \text{HCO}_3^-$.

STANDARD FREE ENERGY OF PEPTIDE BOND SYNTHESIS

Consider the reaction (see equation 109) at 298.1° K:



This gives $\Delta F^\circ = 1.90$ kcal/mole, considerably lower than the value of 4.13 kcal/mole for the synthesis of alanyl-glycine from alanine and glycine, derived from similar data, but decidedly larger than the value of 0.42 kcal/mole for the synthesis of BTGA from benzoyltyrosine and glycylamide (equation 105). It seems doubtful whether so much difference between alanyl-glycine and leucyl-glycine synthesis should be expected. It must be remembered that the values calculated here represent rather small differences between large numbers in the calculations, and careful studies are called for to check the data.

These are merely a few simple illustrations of the uses to which such tables may be put. Extensive illustrations of the use of free energy calculations in biochemistry are given in the article by Borsook (1953) on peptide bond synthesis, and by those of Burton and Krebs (1953), Krebs (1954), and Krebs and Kornberg (1957) on the free energy changes involved in the steps of the tricarboxylic acid cycle and in glycolysis.

GENERAL REFERENCES

General Treatises on Thermodynamics

There are numerous valuable treatises on thermodynamics. We may mention the following:

- Denbigh, K. G. (1955). "The Principles of Chemical Equilibrium," Cambridge University Press, Cambridge. Contains much valuable material, including an introduction to statistical mechanics and related topics.
- Glasstone, S. (1947). "Thermodynamics for Chemists," D. Van Nostrand Co., Princeton, New Jersey. An extremely useful book following a general treatment similar to that of Lewis and Randall, but supplemented by some elementary statistical mechanics and other recent developments. Contains numerous illustrative problems.
- Guggenheim, E. A. (1950). "Thermodynamics. An Advanced Treatment for Chemists and Physicists," Interscience Publishing Co., New York. A valuable but often difficult book; not recommended for a first approach to the subject.

- Klotz, I. M. (1950). "Chemical Thermodynamics," Prentice-Hall, New York. A careful exposition of classical thermodynamics with numerous problems and examples.
- Lewis, G. N., and Randall, M. (1923). "Thermodynamics and the Free Energy of Chemical Substances," McGraw-Hill Book Co., New York. This book remains a classic and is still of great value for the discussion of the application of thermodynamics to chemistry.
- Planck, M. (1930). "Vorlesungen über Thermodynamik," 9th ed., Walter de Gruyter and Co., Berlin. This book is also available in English and had gone through several editions. It remains a great classic on the subject, with lucid exposition of fundamental relations.
- Rossini, F. D. (1950). "Chemical Thermodynamics," John Wiley & Sons, New York, and Chapman and Hall, London.

A Few References on Statistical Mechanics

Of these three books Rushbrooke is the simplest and most elementary; Mayer and Mayer is slightly more advanced, but gives a very clear discussion of the fundamentals as well as many applications. Fowler and Guggenheim is a comprehensive treatise dealing with many important applications of statistical mechanics to physical chemistry.

- Fowler, R. H., and Guggenheim, E. A. (1939; reprinted with corrections 1949). "Statistical Thermodynamics," Cambridge University Press, Cambridge.
- Mayer, J. E., and Mayer, M. G. (1940). "Statistical Mechanics," John Wiley & Sons, New York.
- Rushbrooke, G. S. (1949). "Statistical Mechanics," Clarendon Press, Oxford.

Applications of Thermodynamics to Biochemical Systems

- Borsook, H. (1953). Peptide bond formation, *Advances in Protein Chem.* **8**, 127. Contains a very valuable discussion of the thermodynamics of peptide bond formation.
- Burton, K., and Krebs, H. A. (1953). *Biochem. J.* **54**, 94. A comprehensive survey of free energy data. (Several other papers from the same laboratory which deal with related subjects will be found preceding this in the same volume.)
- Krebs, H. A. (1954). The tricarboxylic acid cycle, in "Chemical Pathways of Metabolism" (D. M. Greenberg, ed.), Vol. I, Chapter 4, Academic Press, New York. Outstanding discussion of the whole cycle with extensive tabulations of free energy data.
- Krebs, H. A., and Kornberg, H. L. (1957). A survey of the energy transformations in living material, *Ergeb. Physiol. biol. Chem. u. exptl. Pharmacol.* **49**, 212. A comprehensive review of major importance. Contains an extensive appendix of free energy data by K. Burton.

OTHER REFERENCES

- Alberty, R. A., Smith, R. M., and Bock, R. M. (1951). *J. Biol. Chem.* **193**, 425.
- Benzinger, T. H., and Hems, R. (1956). *Proc. Natl. Acad. Sci. (U.S.)* **42**, 896; see also T. H. Benzinger, *ibid.* **42**, 109, for a theoretical discussion of this very elegant thermal method.
- Borsook, H. (1953). In "Chemical Pathways of Metabolism" (D. M. Greenberg, ed.), Vol. 2, p. 173, Academic Press, New York.
- Cohn, E. J., and Edsall, J. T. (1943). "Proteins, Amino Acids and Peptides," Reinhold Publishing Corp., New York.

- Dawes, E. A. (1956). "Quantitative Problems in Biochemistry," E. and S. Livingstone, Edinburgh and London. Contains numerous useful problems, taken from the original literature.
- Dobry, A., Fruton, J. S., and Sturtevant, J. M. (1952). *J. Biol. Chem.* **195**, 149.
- Edsall, J. T. (1953). In "The Proteins" (H. Neurath and K. Bailey, eds.), Vol. IB, Chapter 7, Academic Press, New York.
- Flory, P. J. (1953). "Principles of Polymer Chemistry," Cornell University Press, Ithaca, N.Y.
- Gucker, F. T., Jr., Klotz, I. M., and Allen, T. W. (1942). *Chem. Revs.* **30**, 191.
- Hildebrand, J. H., and Scott, R. L. (1950). "Solubility of Non-Electrolytes," Reinhold Publishing Corp., New York.
- Krebs, H. A. (1953). *Biochem. J.* **54**, 78.
- Krebs, H. A., Smyth, D. H., and Evans, E. A., Jr. (1940). *Biochem. J.* **34**, 1041.
- Levintow, L., and Meister, A. (1954). *J. Biol. Chem.* **209**, 265.
- Meyerhof, O., and Green, H. (1948). *J. Biol. Chem.* **178**, 655.
- Pardee, A. B. (1954). In "Chemical Pathways of Metabolism" (D. M. Greenberg, ed.), Vol. 1, Chapter 1, Academic Press, New York.
- Pauling, L. (1940). "The Nature of the Chemical Bond," Cornell University Press, Ithaca, New York, Chapter IX.
- Robbins, E. A., and Boyer, P. D. (1957). *J. Biol. Chem.* **224**, 121.
- Robinson, R. A., and Stokes, R. H. (1955). "Electrolyte Solutions," Academic Press, New York.
- Rushbrooke, G. S. (1949). "Introduction to Statistical Mechanics," Clarendon Press, Oxford.
- Seatchard, G., and Prentiss, S. S. (1934). *J. Am. Chem. Soc.* **56**, 1486, 2314.
- Seatchard, G. (1936). *Chem. Revs.* **19**, 309.
- Scott, E. M., and Powell, R. (1948). *J. Am. Chem. Soc.* **70**, 1104.
- Smith, P. K., and Smith, E. R. B. (1937). *J. Biol. Chem.* **117**, 209; **121**, 607.
- Smith, P. K., and Smith, E. R. B. (1940). *J. Biol. Chem.* **132**, 47, 57.
- Smith, R. M., and Alberty, R. A. (1956a). *J. Phys. Chem.* **60**, 180.
- Smith, R. M., and Alberty, R. A. (1956b). *J. Am. Chem. Soc.* **78**, 2376.

Chapter 5

Electrostatics: Its Application to Polar Molecules and Ionic Solutions

The systems encountered by the biochemist consist largely of highly polar molecules and of ions carrying a net electric charge. Simple cations, such as Na^+ , K^+ , Ca^{++} , and Mg^{++} , are universally present in living cells, as are simple anions such as Cl^- and HCO_3^- . Numerous important intermediates in metabolic reactions, such as adenosine triphosphate, creatine phosphate, and the hexosephosphates, fumarate, malate, oxalacetate, and acetoacetate, under physiological conditions are ions carrying net charges located at one or more points in a rather elaborate chemical structure. The amino acids, peptides, and proteins have already been discussed in some detail in another chapter. These molecules, even when their net charge is zero, are extremely polar, for they are dipolar ions containing one or more pairs of ionic groups of opposite charge, attached to a more or less rigid molecular structure, in which the charges are separated from one another by a considerable distance. At $p\text{H}$ values far from their isoelectric points, all these molecules, of course, are ions carrying a net charge, positive in acid, negative in alkaline solution. Water, the predominant component of the living organism is, as we have already seen, a highly polar molecule. The unusual properties of water are largely due to the polar nature of the molecules and the hydrogen bonds which they form, resulting in their strong attractions for one another and for other ionic and dipolar molecules.

Electrically charged structures give rise to forces of great magnitude, which decrease with increasing distance much more slowly than ordinary intermolecular forces between nonpolar or slightly polar molecules. This is particularly true for ions, but it is also true in large measure for dipolar ions. Biochemical systems and their interactions cannot be properly understood without considering the operation of these forces, which can be described by the laws of classical electrostatics. We, therefore, turn now to a consideration of some of the fundamentals of these laws.

We may take as the starting point for our discussion of electrostatics the law of Coulomb for the force between charged bodies. Consider two small charged bodies in a nonconducting medium, their charges being, respectively, q_1 and q_2 . If they are separated by a distance r , which is very

large compared to the dimensions of either body, then the force, F , between them is given by the equation

$$F = q_1 q_2 / Dr^2 \quad (1)$$

If q_1 and q_2 are of the same sign, the force is, of course, repulsive; if they are of opposite sign, it is attractive. The factor D in equation (1) depends on the medium in which the charged bodies are immersed and is known as the dielectric constant; D is by definition taken as unity if the charged bodies are in a vacuum, but experimentally it is found to be always greater than unity in any other medium. It may thus be defined as the ratio of the forces between two charges of specified magnitude, a given distance apart, if the forces are measured first in a vacuum and then in the medium of dielectric constant D . In terms of this definition, D is a pure number and dimensionless. Other definitions of the dielectric constant may also be given, usually in terms of the capacity of a condenser. If the capacity of the condenser is C_0 with a vacuum between the plates, and C_m with a medium of dielectric constant D filling the space between the plates, then $D = C_m/C_0$. Indeed it is by this method that D is usually measured experimentally. Here, also, it is clear that D is a ratio—in this case, a ratio of two capacities.

For the present, we consider the dielectric constant, D , simply as a parameter characterizing the medium in which the charges are immersed, since many important conclusions can be deduced from this simple picture. It is obvious, however, that D must bear an intimate relation to the structure of the molecules that compose the medium. Since all molecules are made up of electric charges, we may expect that D is a function of the distribution of these charges, and of the effects of the applied electric field in altering this distribution. In Chapter 6, these matters will be considered in more detail. In the meantime we proceed with the use of our very simple model, a set of bodies carrying charges, which are immersed in a hypothetical continuous medium characterized by the value of D . In spite of its limitations, we shall find that many facts can be well explained by this oversimplified model.

The statement of Coulomb's law leads directly to the definition of the electrostatic unit of charge. If two charged bodies carrying charges which are equal and opposite are placed 1 centimeter apart in a vacuum, and if it is found that they attract one another with a force of 1 dyne, then by definition each body carries 1 electrostatic unit of charge, positive in one case, negative in the other. The fact that the charges are equal and opposite may be demonstrated by bringing the two charged bodies into contact and showing that the net charge is then abolished and that electrostatic effects disappear. It should be noted that Coulomb's law in its

simple form (equation 1) holds only if the charged bodies are very small in comparison with the distance between them; we may then, without significant error, think of the charges as located at the center of each charged body, and treat the bodies as if they were points.

All electric charges are exact multiples of the fundamental unit of charge, ϵ , $+\epsilon$ being the charge on the proton, $-\epsilon$ the charge of the electron. The beautiful oil drop experiments of R. A. Millikan showed conclusively by observation of the motion of small oil drops in an electric field that the charge on all electrons is identical to a very high degree of precision. Millikan observed the motion of his oil drops in a combined gravitational and electric field. When a drop acquired or released an electron, its motion in this field of force was suddenly altered, and the magnitude of the charge lost or gained could be calculated from the resulting change of velocity. It was found that this change always corresponded to exactly the same increment or decrement of charge, or to an exact multiple of this fundamental unit, $-\epsilon$, which is thus an exact value, the same for all electrons to within the highest degree of precision of attainable measurements. This fundamental unit of charge, ϵ , is equal to 4.803×10^{-10} electrostatic unit. For most purposes in ordinary calculations, the figure 4.8×10^{-10} is sufficiently precise.

We shall denote vector quantities by symbols in bold-face type (for instance, \mathbf{E}). When we are concerned only with the numerical magnitude of a vector, or sometimes when its orientation is already specified as being in a particular direction, we shall use the corresponding symbol for a scalar quantity, such as E .

Definition of the Potential

Consider a region in space containing one or more charged bodies, which give rise to an electric field. This field determines the magnitude and direction of the force acting on any other charge introduced into the region; it is numerically expressed by the electric intensity, \mathbf{E} , which is a vector quantity denoting the magnitude and direction of the force acting on a unit charge. The electrostatic potential at any point is given by the work required to bring a unit positive charge from an infinite distance, where all electrostatic forces are assumed to be zero, up to the point in question. It is of fundamental importance that this work is independent of the path that is followed by the unit charge in the process of bringing it up to this point; the total net work done is quite independent of the path and depends only on the way the charges are distributed in the system. It is indeed because of this fact that the concept of a potential is valid for an electrostatic field.

The simplest case to consider is that of the potential in the neighbor-

hood of a single charged body which is small enough to be considered as a point charge. Let the charge on this body be denoted by q_1 , and assume it to be immersed in a uniform medium of a dielectric constant D . Then the electric intensity at distance r from q_1 is $E = q_1/Dr^2$. The work of bringing up the charge from $r = \infty$ to a specified finite distance, R , from the charge q_1 is given by the integral of the force acting throughout this distance.

$$W = \int_{r=\infty}^{r=R} -E \, dr = \frac{q_1}{D} \int_{r=\infty}^{r=R} \frac{-dr}{r^2} = \frac{q_1}{DR} \quad (2)$$

The minus signs appear in the integrals in equation (2), since r is decreasing as work is done (W positive) in bringing up the charge. If the system contains n charges distributed at various points in space—say a charge q_1 at a distance R_1 , a charge q_2 at a distance R_2 , from the point in question, etc.—then the potential at the given point is simply the sum of a series of n terms, each corresponding to the single term of equation (2). Thus, denoting the potential by the symbol ψ , we may write:

$$\psi = \sum_{i=1}^n \frac{q_i}{DR_i} \quad (3)$$

The potential is a scalar quantity which is defined, at each point in space, by equation (3). For any given value of the potential, it is found in general that there are regions, generally surfaces, for which all points in the surface have the same value of ψ . Such a surface is known as an equipotential surface. For instance, in a region containing a single point charge, any spherical surface of any radius r , with the charge at its center, is an equipotential surface. Consider any point on an equipotential surface, and suppose that we draw an infinitesimal vector of length dr , in any direction away from this point. If dr lies on the surface, then ψ remains unchanged as we move in space along the vector dr —that is, $d\psi/dr = 0$. This is indeed simply one way of defining an equipotential surface. If dr is normal to the surface, however, then it may be shown quite rigorously that $d\psi/dr$, measured in this direction, has the maximum value possible for any orientation of the vector dr at this point. This maximum value of $d\psi/dr$ is sometimes denoted as $\text{grad } \psi$ or $\nabla\psi$ (where the symbol ∇ is pronounced “del”). If ψ is defined as a function of Cartesian coordinates x, y, z of the points in the system, and if we use the symbols \mathbf{i}, \mathbf{j} , and \mathbf{k} to denote unit vectors parallel to the x, y , and z axes, respectively, then the vector $\nabla\psi$ is given by the equation

$$\nabla\psi = \mathbf{i} \frac{\partial\psi}{\partial x} + \mathbf{j} \frac{\partial\psi}{\partial y} + \mathbf{k} \frac{\partial\psi}{\partial z} \quad (4)$$

Therefore, to obtain the gradient of ψ , we take a vector along the x axis of length $\partial\psi/\partial x$, one along the y axis of length $\partial\psi/\partial y$, and one along the z axis of length $\partial\psi/\partial z$, and add them by the usual rules of vector addition. (Note that each vector must be given its correct sign; thus, if $d\psi/dx$ is negative, $i(\partial\psi/\partial x)$ is a vector pointing along the negative x axis.) It may be readily shown that the value of $\nabla\psi$ is independent of the origin of coordinates chosen, or of any rotation of the axes about this origin. This must hold true for any proper vector, since the value of the vector is

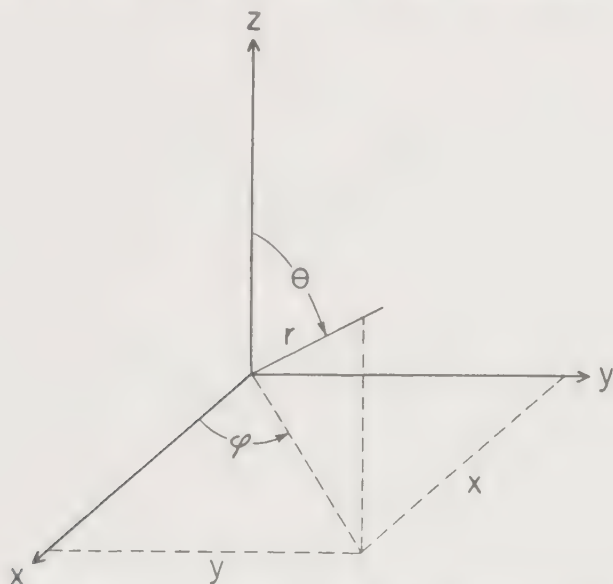


FIG. 1. Relation of Cartesian and polar coordinates.

determined by the physical situation and must obviously be independent of the arbitrary choice of axes.

The negative of the potential gradient gives the electric intensity, \mathbf{E} :

$$\mathbf{E} = -\nabla\psi \quad (5)$$

For instance, in the simple case of the potential at a distance r from a point charge q_1 , it is apparent from symmetry that the equipotential surface is a sphere of radius r , and the potential gradient is normal to this surface. Then, since $\psi = q_1/Dr$, we have $E = -d\psi/dr = q_1/Dr^2$ which is the familiar result. Here we have expressed ψ in spherical polar coordinates, and in this particular instance, the radial distance, r , is the only important coordinate, since ψ is obviously independent of the value of θ , the polar angle, and ϕ , the azimuth.

The equation for the gravitational force between masses is of the same form as Coulomb's law, and such forces are, therefore, derivable from a gravitational potential; the mathematical relations are identical. Physically, however, there are several important differences: (1) the magni-

tude of gravitational attraction for ordinary masses that can be manipulated by the experimenter is very small compared to the electrostatic forces observable between ordinary charged bodies. Indeed for nearly all experimental purposes, gravitational forces are negligible except in astronomy; (2) gravitational forces are only attractive, never repulsive; (3) the gravitational force arising from the interaction between two masses is independent of the nature of the intervening medium, so far as is known. There is thus no factor in gravitation corresponding to the dielectric constant in electrostatics.

Gauss's Law

Electrically charged bodies are commonly too large to be considered as point charges. It is important, therefore, to derive relations from which we may calculate the potential and the laws of force in the neighborhood of charged bodies of finite dimensions. A fundamental relation, derived by Gauss and known by his name, provides a basis for many such calculations.

In the statement of this law, we employ a new concept—that of the normal flux of electric intensity across a surface. If the surface is flat, of area S , and if the electric intensity, E , is normal to the surface and the same at all points, then the normal flux, N , is simply equal to ES . If E makes an angle, θ , with the normal, then $N = ES \cos \theta$. If the surface is curved, we may imagine it divided up into a large number of very small elements, each being so small that it may be taken as flat with negligible error. If dS be the area of one of these elements, and if E in this small region makes an angle θ with the normal to dS , then the contribution of this element to the flux of intensity is $dN = E \cos \theta dS$. It should be noted that this flux is a scalar quantity; we must know the direction of E in order to determine θ , but it is only the magnitude of E that enters the expression $ES \cos \theta$.

It should be clearly understood that the surfaces under discussion here are essentially mathematical constructions; they do not ordinarily correspond to physical boundaries between different phases.

For closed surfaces, to which Gauss's law applies, we adopt the convention that the outer normal to the surface is taken as positive, the inner normal as negative. Consider a closed surface, S , containing a total charge, q , inside it, in a medium of dielectric constant D , which also pervades the medium surrounding S . Then Gauss's law states that the total flux of electric intensity, N , across the surface is equal to $4\pi q/D$.

$$N = \int_S E \cos \theta dS = 4\pi q/D \quad (6)$$

Only charges inside the surface S contribute to N .

This law involves no requirement as to the location of the charge q inside the surface; it may be divided up into smaller charges, and these may be distributed in space in any way we please, so long as they are inside S ; the result, given by equation (6), is the same.

Before giving the general proof of equation (6), consider a particularly simple case—that of a single point charge, q , at the center of a spherical surface in space, of radius r . From symmetry it is clear that E is directed

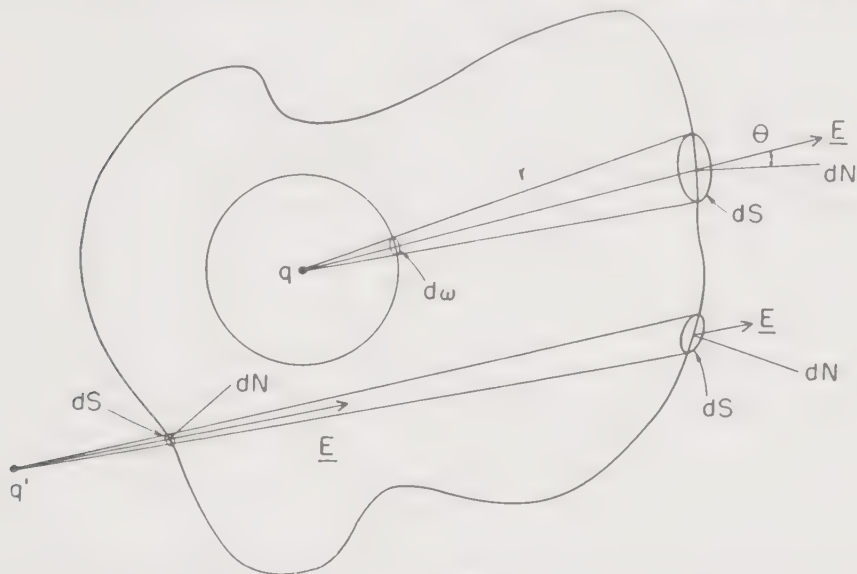


FIG. 2. An illustration of Gauss's law. The irregular closed outline indicates a section of an arbitrary surface in space. The circle shown around the charge q represents a section of a sphere of unit radius, to illustrate the meaning of the element of solid angle $d\omega$. E denotes the electric intensity due to q , dS denotes a surface element of the arbitrary surface, dN denotes the normal to dS , and θ is the angle between E and dN . The charge q' lies *outside* the arbitrary surface; see text.

outward, normal to the surface of the sphere at any point, and is uniformly equal to $E = q/Dr^2$ at all points. Clearly $\theta = 0$ and $\cos \theta = 1$ everywhere. The area of the surface of the sphere is $4\pi r^2$, and the flux of electric intensity is thus $(q/Dr^2) \times (4\pi r^2) = 4\pi q/D$, which verifies Gauss's law for this particular case.

In the more general case, we consider a surface, S , of arbitrary shape, one planar section of which is shown in Fig. 2. Consider a charge, q , anywhere inside the surface, and a small region, dS , of the surface in an arbitrary direction from q , as indicated. Let the distance from q to dS be r (the size of dS being so small that all points within this very small region may be considered to have the same value of r , without sensible error). Then the value of E at dS is q/Dr^2 , and the contribution of the region dS to the total flux of intensity is $dN = q \cos \theta dS/Dr^2$. The solid angle, $d\omega$, which is subtended by dS on a circle of radius r drawn around q as

center, is $d\omega = dS \cos \theta / r^2$.¹ This follows, since the normal to $d\omega$ is drawn directly from q through the center of dS ; it is thus parallel to E . Hence, we obtain

$$dN = q \cos \theta \, dS / Dr^2 = (q/D) \, d\omega \quad (7)$$

This value is the same for any given value of $d\omega$, regardless of the direction of the normal drawn outward from q to the surface, S . Thus, if we integrate over all orientations, the contribution of any region of solid angle ω is simply equal to $(q/D)\omega$, and the total contribution is thus

$$N = (q/D) \int_{\text{all angles}} d\omega = 4\pi q/D \quad (8)$$

which is Gauss's law. If charges q_1, q_2, \dots, q_n are distributed in space within the surface S , any one charge, q_i , contributes the term $4\pi q_i/D$ to N , and the total value of N is still given by equation (8), with

$$q = q_1 + q_2 + \dots + q_n$$

On the other hand, a charge q' outside S (Fig. 2) makes no net contribution to N , for a line (or a small conical pencil of lines) in any given direction cuts dS in two places, if at all. The negative contribution to N , at the point where the line drawn outward from q' enters S , is exactly canceled by the positive contribution at the point where it leaves S .

If we consider surfaces of highly irregular shapes, with many knobs and hollows, a straight line drawn outward from a given point may intersect the surface in several different places. If the point is inside the surface, however, the number of intersections is always odd; if it is outside the surface, always even. Thus a simple extension of the argument just given always leads to the same result, namely equation (8), where q is the total charge inside the surface.

Gauss's law is a direct consequence of the inverse square law of attraction; E varies inversely as r^2 , and dS —for a given value of θ and a given element of solid angle—varies directly as r^2 . The product $E \, dS \cos \theta$ is, therefore, independent of r and depends only on q and D .

One simple and important application of Gauss's law is the calculation of the electric intensity outside any uniformly charged sphere, or spherical shell, of total charge q surrounded by a uniform medium, of dielectric constant D . If the radius of the sphere is b , then we may describe outside of it a concentric sphere of radius $r > b$. From symmetry, it is apparent that the electric intensity must be numerically the same

¹ The solid angle occupied by a surface element dS on a sphere of unit radius (total surface area 4π) is numerically equal to dS ; on a sphere of radius r (total surface area $4\pi r^2$) the solid angle occupied by dS is dS/r^2 .

everywhere on the surface of the outer sphere, and must everywhere be perpendicular to this surface. Hence $N = ES = 4\pi r^2 E = 4\pi q/D$. Thus

$$E = q/Dr^2 \text{ (for spherical symmetry)} \quad (8.1)$$

In other words, the electric intensity outside the charged sphere is the same as if all the charge were concentrated at its center.

Poisson's Equation for Regions Containing a Space Charge

In some problems we are concerned with regions containing a very large number of small charged bodies, so small and so numerous that we may consider the charge as distributed continuously throughout the region. A solution containing positive and negative ions is an important example, which we shall consider later in more detail. If a region of volume ΔV contains n_+ cations of valence Z_+ , and n_- anions of valence Z_- , then the space charge $\rho = e(n_+Z_+ + n_-Z_-)/\Delta V$. (Observe that Z_- is a negative quantity.) In any large region of an ionic solution, electro-neutrality must prevail, and the average value of ρ over such a region is, therefore, zero. In local regions, however, taken around a particular ion as center, there is an unequal distribution of the surrounding ions, owing to electrostatic forces. This fact is indeed the basis of the Debye-Hückel theory, which we shall consider later in detail.

Another type of problem in which the concept of space charge proves useful is in the study of vacuum tubes, in which electrons, emitted from the hot filament, flow through space toward the plate under the influence of the applied field. Here, of course, the problem is one of the flow of current, and outside the domain of electrostatics.

An important electrostatic case is that in which the space charge, ρ , is distributed with spherical symmetry about a center, so that ρ is a function only of r , the distance from the center, and is independent of the angles θ and ϕ , which describe the orientation of the system (Fig. 1). In this case we therefore write $\rho = \rho(r)$, to emphasize this symmetry of charge distribution.

In order to apply Gauss's law to the system we note that both ψ and E are, like $\rho(r)$, symmetrical about the origin, and therefore are functions of r only. This means that, if we describe a sphere of radius r about the origin, the vector \mathbf{E} is everywhere perpendicular to the surface of the sphere and has the same numerical value at every point on the surface. If q is the total charge inside such a sphere, it follows therefore from Gauss's law that

$$\int Er^2 d\omega = 4\pi r^2 E = \frac{4\pi q}{D} \quad (9)$$

The integral on the left of equation (9) is, of course, taken over the surface of the sphere of radius r . Since E is a numerical constant for every point on the surface, regardless of orientation, we have denoted it as a scalar quantity in this equation. Now, remembering that $E = -d\psi/dr$,

$$r^2 \frac{d\psi}{dr} = - \frac{q}{D} \quad (10)$$

At the same time the infinitesimal amount of charge dq contained in any spherical shell of radius r and thickness dr with its center at the origin is

$$dq = 4\pi r^2 \rho(r) dr \quad (11)$$

from which it follows that

$$\frac{dq}{dr} = 4\pi r^2 \rho(r) \quad (12)$$

By differentiating (10) with respect to r , combining with (12), and dividing both sides by r^2 , we obtain

$$\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{d\psi}{dr} \right) = - \frac{4\pi \rho(r)}{D} \quad (13)$$

This is Poisson's equation, relating the density of space charge to the variation of the potential gradient with the coordinates, for the special case of a system with spherical symmetry.

We may, of course, write out the derivative on the left-hand side of (13), in which case Poisson's equation becomes

$$\frac{2d\psi}{r dr} + \frac{d^2\psi}{dr^2} = - \frac{4\pi}{D} \rho(r) \quad (13.1)$$

A special type of spherically symmetrical system, which we shall encounter in considering the Debye-Hückel theory, has at its center a spherical ion of radius b and charge $Z\epsilon$. This charge is considered to be distributed uniformly over the surface of the ion, as if the latter were a small conducting sphere. Other charges can approach the center of this ion only to the distance a , which is greater than b . The whole region outside the central ion ($r > b$) is considered to have a constant dielectric constant, D . The charge density, ρ , is zero between b and a ; at a suddenly jumps to a finite value; at values of $r > a$, ρ is a continuous function of r , which remains always finite or zero. The total charge enclosed within any sphere of radius between b and a is simply the charge ($Z\epsilon$) on the central ion, since there is no space charge in this region; hence

from Gauss's law we have (compare equation 9) $r^2E = Z\epsilon/D$, as in the simple case of the charged sphere already discussed on p. 248. At values of r greater than a we have, for the total charge, q , enclosed by a sphere of radius r ,

$$q = Z\epsilon + \int_a^r 4\pi r^2 \rho(r) dr$$

Hence, by the same reasoning used in equation (9), we may write for any value of r greater than a

$$r^2E = \frac{q}{D} = \frac{Z\epsilon}{D} + \frac{4\pi}{D} \int_a^r r^2 \rho(r) dr \quad (13.2)$$

The value of the integral in (13.2) approaches zero as r approaches a from above, since ρ is everywhere finite (or zero). Thus, although the value of ρ changes discontinuously from zero to a finite value at $r = a$, E is a continuous function of r everywhere from the surface of the central ion ($r = b$) outward. The same is therefore also true of the potential

$$\psi = - \int_{\infty}^r E dr$$

On differentiating (13.2) with respect to r , it is obvious that the derivative of the constant term $Z\epsilon$ is zero; the derivative of the integral immediately gives equation (12) again, from which Poisson's equation (13) follows. In the discussion of the Debye-Hückel theory we shall see how $\rho(r)$ can be evaluated for a model system which corresponds closely to many actual systems containing ions.

This discussion has been confined to the special case of spherical symmetry. We may remark, without proof, that the general form of Poisson's equation in Cartesian coordinates is

$$\nabla^2 \psi = \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} = - \frac{4\pi\rho}{D} \quad (14)$$

with a corresponding expression in spherical polar coordinates. The symbol ∇^2 (read as "del squared") is the Laplacian operator which, when applied to a scalar function of the coordinates such as a potential, yields the expression given above. In regions where the charge density is zero, the right-hand side of (14) vanishes, and we obtain Laplace's equation:

$$\nabla^2 \psi = 0 \quad (15)$$

which is of central importance, not only in electrical theory, but in many other branches of mathematical physics.

Free Energy of a Charged Sphere

Simple ions, such as Na^+ , Ca^{++} , or Cl^- , may be expected to possess spherical symmetry; except for the fact that they carry a net charge, their electronic structure is like that of the inert gases, which are—at least in their average configuration—spherical in shape. If we consider an ionic solution—for instance, a solution containing the ions Na^+ and Cl^- —we may imagine a spherical cation and anion being formed from a corresponding pair of neutral spheres. This involves, of course, the removal of a charge from one of the spheres and its transference to the other. The sphere from which the charge has been removed thus carries a charge of the opposite sign, and the whole process requires work in order to separate the charges. The process of separating the charges is reversible, if it is carried out sufficiently slowly so that forces analogous to friction are reduced to negligible proportions. Thus the work of charging the spheres represents an increment in the free energy of the system; we may designate this as electrical free energy. Qualitatively it is apparent that the work of charging, and hence the electrical free energy, increases as the dielectric constant of the medium decreases.

In order to calculate this work of charging in quantitative terms, we make use of a model derived from classical electrostatics, that of a uniform conducting sphere, the charge on which can be altered from zero up to any desired value by adding successive infinitesimal increments of charge, the sphere of radius b being immersed in a medium of dielectric constant D . We start with the sphere uncharged, and bring up small increments of charge, until the final charge on the sphere is q . We may denote the charge at an intermediate state in the process by λq , where λ is a fraction varying from zero (when the sphere is uncharged) to unity (when the process is complete). Each successive small increment of charge brought up increases λq by a small amount, and we may denote this increment by $d(\lambda q)$. The electric intensity outside the sphere is given by (8.1) as $\lambda q / Dr^2$, if $r > b$. The potential at the surface of the sphere is, therefore, $\psi = \lambda q / Db$, and the work increment, dW , involved in bringing up the additional charge increment $d(\lambda q)$ is

$$dW = \psi d(\lambda q) = \lambda q d(\lambda q) / Db \quad (16)$$

The total work, W , of charging the sphere from zero to q is then given by integrating this expression, noting that q , the final charge, is a constant and that the only variable is λ , which varies from zero to unity:

$$W = \int_{\lambda=0}^{\lambda=1} \lambda q d(\lambda q) / Db = q^2 / Db \int_{\lambda=0}^{\lambda=1} \lambda d\lambda = q^2 / 2Db \quad (17)$$

The electrical free energy of the sphere, which is equal to W , is thus greater, the smaller the radius of the sphere and the smaller the dielectric constant. It is proportional to the square of the charge on the sphere, and thus is independent of the sign of charge.

The work increment, dW , is quite generally equal to the potential at the surface of the sphere, ψ , multiplied by the increment of charge brought up to the surface. The formula $\psi = \lambda q/Db$, however, applies only to an isolated sphere, and the potential so evaluated is sometimes called the self-potential of the sphere. If there are other charged bodies in the neighborhood, the field due to them affects the potential at the surface of the sphere in which we are interested, and, therefore, affects its electrical free energy. The effect of a charge q_i at a distance r_i from the sphere—assuming that $r_i \gg b$ —on the potential of the sphere is simply q_i/Dr_i , and the effect of a collection of such charges distributed in space around the sphere is simply the sum of the separate terms, $\sum_i (q_i/Dr_i)$.

Note that each q_i may be either positive or negative; so if the central sphere is surrounded by many positive and many negative charges, their net effects may largely cancel.

If we have a collection of charged spheres, each in a specified position and all immersed in a uniform medium of dielectric constant D , we may calculate the electrical free energy of the whole assembly of spheres by an argument similar to that already given for a single sphere. We begin with all the spheres in position, but uncharged, and bring up to each sphere in the system a small element of charge, $d(\lambda q_i)$, which is proportional to the total final charge which the sphere in question is to receive. Thus, at any stage in the charging process, the charge on the i th sphere is λq_i , where λ —again a fraction varying between zero and unity—is the same fraction for all the spheres in the system. Thus the potential at the surface of the i th sphere is equal to $\lambda\psi_i$, if ψ_i is its final potential when charging is complete. Thus the work of charging the i th sphere is given by the expression

$$W_i = \int_{\lambda=0}^{\lambda=1} \lambda\psi_i d(\lambda q_i) = \psi_i q_i/2 = (F_e)_i \quad (18)$$

Thus the electrical free energy of each sphere, $(F_e)_i$, is equal to half its charge multiplied by its potential, a relation which is also true in the more special case of equation (17). Moreover, for a system of charged spheres separated by distances which are large compared to the radius of any one of them, we may write, by the argument given above,

$$\psi_i = q_i/Db_i + \sum (q_j/Dr_{ij})$$

where the summation over the j 's includes all spheres except the i th sphere, and r_{ij} is the distance from the i th to the j th sphere.

The Potential and Energy of a Dipole

Any atom or molecule consists of an aggregation of charges—positive charges associated with atomic nuclei and negative charges due to surrounding electron shells—and most molecules, even in the absence of a net charge, are the seat of an appreciable electric field, arising from local concentrations of charge within them. The higher the symmetry of the molecule, of course, the smaller in general will be the effect, and in a molecule like benzene it may, for most purposes of the chemist, be neglected.

As a first approximation the field surrounding a molecule which has no net charge may be analyzed in terms of what is known as the dipole moment of the molecule. In evaluating this moment, we make use of the concept of the center of positive or negative charge within the molecule. This is completely analogous to the concept of the center of mass of a system of particles. Except in molecules like benzene and carbon tetrachloride, having a very high degree of symmetry, and in some other special cases such as various hydrocarbons, we must expect that there will be a lack of coincidence between the centers of the two kinds of charges, and where this is the case, the product of the absolute magnitude of either charge (they will, of course, be equal in an uncharged molecule) times the distance between the centers gives the dipole moment. Some idea of the magnitude of the effect to be expected is given by the moment of an electron and a proton separated in a fixed configuration by a distance of 1 Å. This is equal to $4.8 \times 10^{-10} \times 10^{-8} = 4.8 \times 10^{-18}$ electrostatic unit. This is of the order of the moments of many small molecules, though larger than most. For example: the moment of the water molecule is about 1.9×10^{-18} .

The significance of this concept and the reason for the name will be clear from the following considerations. If a molecule possessing a dipole moment is placed in a uniform electric field, the positive charges will be subject to a net force acting in one direction at *their* center and the negative charges to a net force acting in the opposite direction at *their* center. The molecule as a whole will be subject to a torque tending to line it up in the direction of the field, and the magnitude of this torque will be the product of the magnitude of either force times the perpendicular distance between them. Suppose the strength of the field is E , the absolute magnitude of positive or negative charge is q , the distance between the centers is l , and the angle between l and E is θ . Then the magnitude of the force acting on either charge is qE , and the torque acting on the molecule as a whole is (see Fig. 3).

$$T = Eql \sin \theta \quad (19)$$

If we denote the dipole moment by

$$\mu = ql \quad (20)$$

this becomes

$$T = E\mu \sin \theta \quad (21)$$

For many purposes, it is convenient to treat μ as a vector having the direction of l , in which case we write $\mathbf{u} = ql$. In terms of μ , equation (21) may be expressed in vector notation as

$$\mathbf{T} = \mathbf{u} \times \mathbf{E} \quad (22)$$

when $\mathbf{u} \times \mathbf{E}$ is the vector product of \mathbf{u} and \mathbf{E} .²

It should be realized that under the influence of an external electric field any molecule will develop a dipole moment (also called an electric

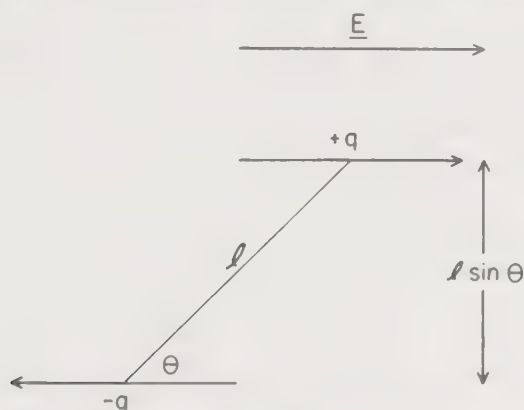
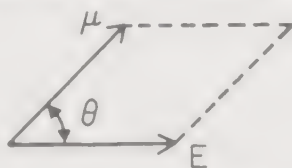


FIG. 3. Torque on a dipole due to an electric field of intensity \mathbf{E} .

moment) due to the displacement of charge within it produced by the field. Indeed this is what is responsible for the ordinary optical index of refraction of a molecule, light being an electromagnetic wave. But certain molecules—in fact, as we have pointed out, all but a few—will have an inherent dipole moment which results from their native structure, quite apart from any distortion effects. Such dipole moments are called per-

² By definition the vector product $\mathbf{T} = \mathbf{u} \times \mathbf{E}$ has the *magnitude* $\mu E \sin \theta$, where θ is the angle between the two vectors. This magnitude is given by the area of the parallelogram in the accompanying sketch. The *direction* of \mathbf{T} is perpendicular to \mathbf{u}



and \mathbf{E} , such that the shortest rotation of \mathbf{u} which brings it into coincidence with \mathbf{E} is, as seen from the positive end of \mathbf{T} , counterclockwise. Thus, if \mathbf{u} and \mathbf{E} lie in the plane of the paper as shown, \mathbf{T} is directed backward, away from the reader.

manent dipole moments, to distinguish them from induced moments, and the molecules which possess them are called permanent dipoles. They are also spoken of as polar molecules.

Let us now consider the field due to such a dipole. This is best described in terms of its potential, ψ , which may be expressed as the sum of two potentials, one for each of the two charges of the dipole. At any point

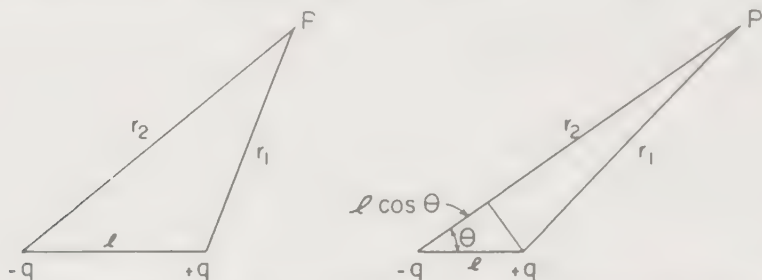


FIG. 4. The potential at a point P due to a dipole with charges $-q$ and $+q$ at the points indicated. The drawing on the right indicates the condition that is approached as the distance l between the dipole charges approaches zero.

P at a distance r_1 from the positive charge and r_2 from the negative charge, the potential is given by

$$\psi = \left(\frac{q}{r_1} - \frac{q}{r_2} \right) \frac{1}{D}$$

This is clearly symmetrical about the vector \mathbf{l} , joining the charges, which is called the dipole axis. If desired, r_2 may be expressed in terms of r and θ , the latter being the angle between r_1 and the dipole axis (see Fig. 4):

$$r_2^2 = r_1^2 + l^2 + 2r_1l \cos \theta$$

For values of r_1 large in comparison with l , r_2 is given with greater or less approximation, depending on the ratio l/r_1 , by

$$r_2 = r_1 + l \cos \theta$$

In this case the expression for ψ becomes

$$\psi = q \left(\frac{1}{r_1} - \frac{1}{r_1 + l \cos \theta} \right) \frac{1}{D} \cong \frac{ql}{r_1^2 D} \cos \theta \cong \frac{\mu \cos \theta}{r_1^2 D} \quad (23)$$

Under these conditions, it is unnecessary to distinguish between r_1 and r_2 , the ratio of which approaches unity, and we may, therefore, drop the subscript 1, identifying r simply with the distance of the dipole from the point P , and θ with the angle between the dipole axis and the line joining the dipole to P . In this connection it should be recalled that the dipole

axis is to be taken as a vector pointing from the negative to the positive charge of the dipole.

For many purposes, particularly as a means of simplifying mathematical calculations, it is often helpful to make use of the concept of a point dipole. A point dipole is nothing more than the limiting case approached by any actual dipole when the distance between the charges is reduced to zero while at the same time their magnitude is correspondingly increased, so as to keep the product of the two (ql), which gives the dipole moment, constant. A point dipole preserves the vector property of actual dipoles. Although strictly speaking a point dipole must be regarded as a fiction, it represents a useful approximation for many purposes. By treating any actual dipole as if it were a point dipole, it is possible to apply the limiting equation for its potential at all distances.

As we have said, the majority of molecules may be expected to have some permanent dipole moment. The limiting case of nonpolar molecules is represented by molecules having a high degree of symmetry such as benzene and many of the hydrocarbons. The opposite extreme is represented by a class of molecules known as dipolar ions or, in German, *Zwitterions*. These are molecules such as the amino acids in neutral solution, which contain two or more oppositely charged groups resulting from ionization. Glycine or alanine, over a wide range of pH , exist mainly in a form in which at one side of the molecule there is the negatively charged $-\text{COO}^-$ group, and on the other a positively charged $-\text{NH}_3^+$. A more extreme example of a dipolar ion is presented by such a molecule as lysylglutamic acid, which contains two positively and two negatively ionized groups. Larger peptides and proteins present still more remarkable cases. Molecules of this sort are in a class by themselves as regards dipole moments, and in many respects behave accordingly. They are of great interest and importance to the biochemist, particularly since all the proteins fall into this class. The study of dipolar ions is a matter which will occupy us repeatedly in later chapters.

Before concluding this section, it is important to consider the expression for the energy of an actual dipole. Of course this will depend on the exact model which we adopt. A convenient one, which is well suited as a representation of a simple dipolar ion like glycine, consists of two spheres of radius b_1 and b_2 , uniformly charged with charges $+q$ and $-q$, and separated by a distance R measured between centers. In order to reckon the energy of such a dipole, we consider the electrical work of producing it. We may think of the dipole as being formed by first charging the spheres when they are sufficiently far apart so that electrostatic interactions between them may be neglected, and then bringing them to the required separation, R . We suppose the dipole to be immersed in a con-

tinuous medium of dielectric constant D . The work of charging the two spheres is (by 17)

$$W = \frac{q^2}{2D} \left(\frac{1}{b_1} + \frac{1}{b_2} \right) \quad (24)$$

One of these terms is for the positively charged sphere, the other for the negatively charged sphere. If we assume that the uniform distribution of charge on each sphere does not change as the spheres are brought together, the additional work of bringing them together may be easily calculated. Since at any point external to a uniformly charged sphere or shell the field everywhere is the same as if all the charges were located at the center, this work is given by the potential due to either charge at a distance R multiplied by the other charge. This gives $-q^2/RD$. The total work of charging the dipole, and consequently its total energy, is therefore given by

$$W = \frac{q^2}{D} \left(\frac{1}{2b_1} + \frac{1}{2b_2} - \frac{1}{R} \right) \quad (25)$$

If $b_1 = b_2$, corresponding to the case of two equal spheres, this reduces to

$$W = \frac{q^2}{D} \left(\frac{1}{b} - \frac{1}{R} \right) \quad (26)$$

It is sometimes desirable to use other, more refined models for calculations concerning dipolar ions, but this simple one is often useful.

Image Charges

The electrostatic constructions known as image charges are frequently useful in problems relating to molecules in solution. They are employed, for instance, in the theory of the salting-out effect, to calculate interactions between ions and a surrounding dielectric medium. The most natural way of introducing the concept is in connection with the problem of calculating the influence of a conducting sphere, maintained at zero potential by connection with ground, on the field due to a neighboring point charge. In the presence of the point charge, the sphere will, of course, acquire a charge of opposite sign from ground. If the magnitude of this charge and its distribution over the surface of the sphere were known, it would be possible, though perhaps not easy, to calculate the potential at any point outside the sphere. But instead of proceeding in this way, it is far simpler to replace the sphere conceptually by a suitably located ideal point charge, of magnitude equal to the total charge acquired by the sphere, and base the calculations on this ideal point charge. It can be shown that everywhere *outside the region occupied by the sphere*

the effect of the ideal point charge will be the same as that of the sphere. Such an ideal point charge is known as an image charge and may be regarded as the electrical image of the actual external charge. The basis of this will be clear from the following considerations.

Consider two charges, e_A and e_B , of unequal magnitude and opposite sign, located at points A and B in a region where the dielectric constant is D . At any point P , distant r_A from A and r_B from B , the potential ψ is given by

$$\psi = \left(\frac{e_A}{r_A} + \frac{e_B}{r_B} \right) \frac{1}{D} \quad (27)$$

Any given value of ψ defines a surface of constant potential. In general such surfaces will be complicated, being in fact surfaces of the eighth degree. For the special case where $\psi = 0$, however, the surface becomes a sphere surrounding the charge of smaller absolute magnitude, and with

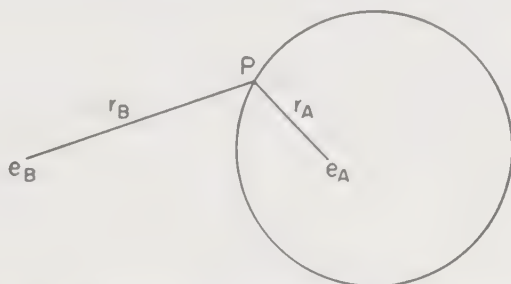


FIG. 5. Equipotential surface ($\psi = 0$) in a field due to the two charges e_A and e_B .

its center on the line joining the charges (see Fig. 5). This can easily be verified by writing the expressions for r_A and r_B in Cartesian coordinates. If, therefore, an actual conducting sphere, maintained at zero potential ($\psi = 0$) and of the same radius as the geometrical sphere defined by $\psi = 0$, is substituted for the latter and the charge which it contains, the potential remains everywhere unchanged *in the region outside the sphere*. This may be taken as self-evident, but a formal proof follows from the theorem that the value of any function which satisfies Laplace's equation (as ψ must of course do) is uniquely determined in any region (in this case the region outside the sphere) by its values at the boundary of the region. The only condition is that the function, together with its first derivative, be continuous.³ The smaller charge, located inside the sphere, is to be regarded as the electrical image of the other charge, produced by the conducting sphere maintained at zero potential. The magnitude of this image charge and its position inside the sphere are uniquely determined

³ We need not go into the complications arising from the singular point occupied by the external point charge where E and ψ both become infinite.

by the radius of the sphere, R , the size of the external charge, which we suppose to be e_B , and the distance of e_B from the center of the sphere, which we denote by b (Fig. 6). Let the distance of e_A from the center of the sphere be a . We have two equations for the determination of the two unknowns, e_A and a . These result from the condition that $\psi = 0$ for the

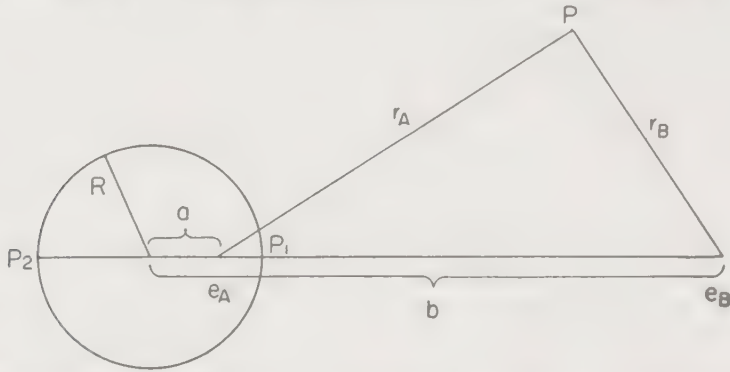


FIG. 6. The determination of the magnitude and position of the image charge e_A .

two points of intersection, P_1 and P_2 , of a straight line through the charges and the surface of the sphere, namely:

$$\frac{e_A}{R + a} = - \frac{e_B}{R + b} \quad (28)$$

and

$$\frac{e_A}{R - a} = \frac{-e_B}{b - R} \quad (29)$$

Solution of these equations gives

$$a = \frac{R^2}{b} \quad (30)$$

and

$$-e_A = e_B \frac{R}{b} \quad (31)$$

Of course the problem could be inverted; e_A could play the role of an actual point charge contained within a spherical cavity in an indefinitely extended conducting medium maintained at zero potential. Then e_B would be the electrical image of e_A produced in the surrounding conductor. There is, therefore, a reciprocal relation between a charge and its image charge.

The concept of an image charge is not limited to conductors but has a much more general applicability. This is illustrated by two further simple examples. The first is that of a point charge, e , situated at the center of a spherical region of dielectric constant D_1 and of radius a . Suppose

that, outside the region ($r > a$), the dielectric constant is everywhere D_2 out to infinity. Consider the potential at any point, r , inside the region. This will be the same as the work required to bring up a unit charge from infinity to the point r . This work may be broken up into two parts, that of bringing the charge from infinity to the surface of the sphere $r = a$, and that of bringing it from a to the point $r < a$. The first occurs in the region where the dielectric constant is D_2 and the electric intensity, by Gauss's law, taking account of spherical symmetry (equation 8.1), is e/D_2r^2 . The second occurs in the region where the dielectric constant is D_1 , and the electric intensity by the same law is e/D_1r^2 . The total work, W , done on the particle is therefore

$$\begin{aligned} W &= - \int_{\infty}^a \frac{e}{D_2r^2} dr - \int_a^r \frac{e}{D_1r^2} dr \\ &= \frac{e}{aD_2} + \frac{e}{D_1} \left(\frac{1}{r} - \frac{1}{a} \right) = \frac{e}{D_1r} + \frac{e}{a} \left(\frac{1}{D_2} - \frac{1}{D_1} \right) \end{aligned} \quad (32)$$

If there were no surrounding medium—that is, if the spherical region of dielectric constant D_1 extended out to infinity—the potential of the point charge would be e/D_1r . The effect of the surrounding medium *on the region inside* is, therefore, to increase the potential everywhere by a constant amount

$$\frac{e}{a} \left(\frac{1}{D_2} - \frac{1}{D_1} \right)$$

This may be interpreted as the effect of a total charge

$$e \left(\frac{1}{D_2} - \frac{1}{D_1} \right) D_1$$

uniformly distributed over the spherical surface of the sphere $r = a$ separating the two regions, assuming the region of dielectric constant D_1 to extend out to infinity. The potential of such a charge distribution is known to be everywhere constant inside the sphere and equal to its value at the surface of the sphere, where it may be reckoned as if the charge were concentrated at the center of the sphere. Depending on whether D_1 is less or greater than D_2 , this charge will be either positive or negative. It may be regarded as an image of the central charge, e , resulting from the presence of the surrounding medium. The whole phenomenon is intimately related to the question of dielectric polarization, which will be dealt with in Chapter 6.

The second example is that of reckoning the effect of an insulating sphere of radius a and dielectric constant D_1 on the field in an indefinitely

extended medium of dielectric constant D_2 . It is assumed that in the absence of the sphere the electric intensity would have everywhere the constant value E_1 , which will be the same as the actual value at any point far removed from the sphere. The solution of this problem involves somewhat more elaborate methods than we have used so far and is given in Chapter 6.⁴ For purposes of illustrating the concept of image charges the result may be anticipated here without proof. It may be stated in terms of a system of spherical coordinates with origin at the center of the sphere and axis antiparallel to the undisturbed electric intensity \mathbf{E}_1 . At any point (r, θ) *outside the sphere* the potential, which for obvious reasons of symmetry is independent of the azimuthal angle ϕ , is given by

$$\psi = E_1 r \cos \theta + a^3 \left(\frac{D_2 - D_1}{2D_2 + D_1} \right) E_1 \frac{\cos \theta}{r^2} \quad (33)$$

The first term in this expression is the undisturbed potential, i.e., the potential as it would be everywhere in the absence of the sphere. The second represents the effect of the sphere. On the basis of concepts developed in the previous section (see equation 23), this second term may be interpreted as the potential of a point dipole of moment

$$\mathbf{u} = -a^3 \left(\frac{D_2 - D_1}{2D_2 + D_1} \right) D_2 \mathbf{E}_1 \quad (33.1)$$

located at the center of the sphere. According to whether D_1 is less or greater than D_2 the axis of this dipole will be either antiparallel or parallel to \mathbf{E}_1 . In either case its moment is proportional to the volume of the sphere. The dipole may be regarded as representing a rather special type of distribution of image charges. It may also be thought of in terms of the polarization of the two media, one inside, the other outside, the sphere.

There are two interesting special cases of this example. The first results from setting $D_1 = 1$. This corresponds to making the sphere an empty cavity. In this case the axis of the dipole is antiparallel to \mathbf{E}_1 and its moment is simply

$$a^3 \left(\frac{D_2 - 1}{2D_2 + 1} \right) D_2 \mathbf{E}_1$$

The dipole is to be interpreted as arising wholly from the polarization of the medium surrounding the cavity. The second case is obtained by letting D_1 increase to infinity. This is equivalent to making the sphere a conductor, for from a formal point of view a conductor may be regarded as a medium of infinite dielectric constant. In this case, the axis of the dipole

⁴ See the section of Chapter 6 entitled "Modification of the Debye Theory."

is parallel to \mathbf{E}_1 . The moment is simply $a^3\mathbf{E}_1$, and it depends on the volume of the sphere and not at all on the properties of the surrounding medium.

It will be realized that a uniform electric intensity E can be thought of as arising from an indefinitely large point charge at an indefinitely great distance from the sphere. As the charge moves up to the sphere, being at the same time correspondingly reduced, the situation changes quantitatively, but not qualitatively. The effect of the image charges will still be represented by a dipole in the sphere, though, of course, the exact equations will become more complicated. A rigorous solution of this problem involves advanced methods and the use of spherical harmonics. But we do not wish to cloud the principles by too much detail. The purpose of this section has been to give an idea of the meaning and usefulness of the concept of image charge, and its relation to other concepts.

The Salting-Out Effect

Simple gases and most slightly soluble organic molecules are less soluble in most salt solutions than in pure water, the logarithm of the solubility being approximately a linear function of the concentration of the added salt. For example, diethyl ether, which dissolves in water at 23° to form a saturated solution at mole fraction 0.0154 (weight fraction 0.0607) is about 40% less soluble in the presence of molar sodium or potassium chloride (data cited by Debye, 1927). In solutions of ethanol in water, addition of salts increases the vapor pressure of the ethanol, while decreasing that of the water. The addition of high concentrations of certain salts containing divalent ions—for example, potassium carbonate—produces so great an effect in this direction that the water-ethanol mixture separates into two phases: a denser phase rich in water and salt, and poor in ethanol, and a lighter phase, rich in ethanol, with relatively little water and salt. From a thermodynamic point of view, all these phenomena are expressions of the fact that the addition of salt increases the activity coefficient of the dissolved gas or organic molecule, while lowering that of the water. The relation is reciprocal—if addition of component B to the system increases the activity coefficient of component A , then it follows of necessity, as is indicated in more detail below, that addition of A increases the activity coefficient of B . Thus addition of most organic molecules to aqueous salt solutions decreases the solubility of the salt; or if the salt is too soluble to be studied readily in this fashion, its activity may be shown to be increased owing to the addition of the organic molecule, by such techniques as electromotive force measurements, using electrodes reversible to the ions of the salt. (Some electrodes of this type are discussed in Chapter 8.)

The whole variegated set of phenomena described above may be designated by a common term—the salting-out effect. This effect is a matter of considerable importance in determining the distribution coefficients of organic molecules between aqueous solutions and other phases immiscible with water; it may thus influence passage of molecules across phase boundaries in biochemical systems. It can also greatly affect fractionation procedures based on partition of molecules between different phases.

Perhaps the most striking examples of the salting-out effect are found in protein solutions; differential precipitation of proteins by the addition of high concentrations of salts such as ammonium sulfate or sodium sulfate has been employed in protein separations for at least a century. The salting-out effect in protein solutions is indeed commonly manifest only at high salt concentrations, of the order of several moles per liter; but in the range in which salting out is effective, relatively small increments of salt concentration produce great decreases in protein solubility.

There are a few notable examples of the opposite effect, which may be called “salting in.” Thus hydrocyanic acid becomes more soluble in water in the presence of neutral salts, and so do many amino acids, such as glycine and cystine. Hydrocyanic acid and glycine are alike in that their aqueous solutions have higher dielectric constants than pure water,⁵ whereas the addition of most organic molecules lowers the dielectric constant of water. Indeed, in a homologous series of molecules, the progressive addition of nonpolar hydrocarbon residues causes a progressive decrement in the dielectric constant of an aqueous solution, per mole of solute; and it also is associated with an increasing tendency for the molecule to be salted out by added salts. These facts suggest that the underlying mechanism of the salting-out effect may be largely electrostatic, depending on the electrical energy of the system in the presence of the intense electrostatic fields around the ions of the salt.

The simplest model system to consider (Debye and Macaulay, 1924; Scatchard, 1927) is to treat each ion in the system as a small charged sphere. If there is only one kind of cation and one kind of anion in the system we may call the radius of the cation b_c and that of the anion b_a . The charge on each ion is, of course, an exact multiple of the proton charge, e . If we assume the valences of both kinds of ion to be the same, as in the case of the NaCl or MgSO₄, and of absolute magnitude Z , the charge on the cation is Ze and that on the anion is $-Ze$. Then, by equation (17), the work of charging the cation, in a medium of dielectric constant D , is $W_c = Z^2e^2/2Db_c$, and that of charging the anion is $W_a = Z^2e^2/2Db_a$.

⁵ Cystine is so insoluble that the dielectric constants of its solutions have never been measured, but they also would almost certainly be higher than that of pure water.

All this work represents added free energy of the system, for the idealized process of charging, as defined in the derivation of (17), is completely reversible. The electrical free energy per mole of cations plus anions is of course N times as great as the sum of the two work terms given above, where N is Avogadro's number. We assume that the concentration of salt is small, so that F_e , the electrical free energy, is proportional to the number of salt ions added. Thus in a system containing n_3 moles of salt the electrical free energy is

$$F_e = \left(\frac{NZ^2\epsilon^2}{2Db_c} + \frac{NZ^2\epsilon^2}{2Db_a} \right) n_3 = \left(\frac{NZ^2\epsilon^2}{b} \right) \frac{n_3}{D} \quad (34)$$

Here we have defined a "mean radius," b , for the ions by the equation

$$\frac{2}{b} = \frac{1}{b_c} + \frac{1}{b_a} \quad (35)$$

since in practice there is no way of distinguishing between the two individual radii. If we denote the dielectric constant of water by D_0 , then the electrical free energy change, ΔF_e , involved in transferring n_3 moles of ions from water (commonly chosen as the standard medium) to another medium of dielectric constant D is

$$\Delta F_e = F_e - F_e^\circ = \frac{NZ^2\epsilon^2}{b} \left(\frac{1}{D} - \frac{1}{D_0} \right) n_3 \quad (36)$$

More generally, if each salt molecule dissociates to give a total of ν ions, ν_+ cations of valence Z_+ and ν_- anions of valence Z_- , the free energy change due to the transfer of n_3 moles of salt, i.e., $n_3\nu_+$ moles of cations and $n_3\nu_-$ moles of anions, becomes

$$\Delta F_e = N\epsilon^2 \left(\frac{\nu_+ Z_+^2}{2b_c} + \frac{\nu_- Z_-^2}{2b_a} \right) \left(\frac{1}{D} - \frac{1}{D_0} \right) n_3$$

or

$$\Delta F_e = A \left(\frac{1}{D} - \frac{1}{D_0} \right) n_3 \quad (37)$$

where A is a constant characteristic of the salt. It may be noted that if we introduce the equations $\nu = \nu_+ + \nu_-$ and $\nu_+ Z_+ + \nu_- Z_- = 0$, the second of which represents electrical neutrality, we may eliminate ν_+ and ν_- from (37), with the result

$$\begin{aligned} \Delta F_e &= - \frac{N\epsilon^2 \nu Z_+ Z_-}{Z_+ - Z_-} \left(\frac{Z_+}{2b_c} - \frac{Z_-}{2b_a} \right) \left(\frac{1}{D} - \frac{1}{D_0} \right) n_3 \\ &= - \frac{N\epsilon^2 \nu Z_+ Z_-}{2b} \left(\frac{1}{D} - \frac{1}{D_0} \right) n_3 \end{aligned} \quad (37.1)$$

where b is a mean radius defined by

$$\frac{1}{b} = \frac{2}{Z_+ - Z_-} \left(\frac{Z_+}{2b_c} - \frac{Z_-}{2b_a} \right)$$

This equation is of the same form as (36) and serves to define A . The terms Z_+ and Z_- are signed quantities; hence $Z_+ - Z_-$ is always positive, and Z_+Z_- is always negative.

We may consider the chemical potential* of the salt (μ_3) as made up of two terms—the electrical term $\mu_{3_e} = \partial F_e / \partial n_3$, and the ideal term, μ_{3_I} , that would describe the behavior of a component of an ideal or perfect solution, as defined in Chapter 4. The latter term is given by the equation

$$\mu_{3_I} - \mu_{3_I}^\circ = RT \ln (N_3/N_3^\circ)$$

Here N_3 is the mole fraction of component 3 (the salt), and N_3° is its mole fraction in a dilute aqueous solution which is taken as a standard of reference. The definition of mole fraction in a system containing a salt that dissociates into ions must be made with care. One mole of salt gives ν moles of ions. Thus, in a system containing n_1 moles of water, n_2 moles of a second uncharged component, such as ethanol or ether, and n_3 moles of salt, the total number of moles or particles (molecules and ions alike being counted as particles) is $n_1 + n_2 + \nu n_3$. Therefore we define the mole fraction of salt, N_3 , by the equation

$$N_3 = \frac{\nu n_3}{n_1 + n_2 + \nu n_3}$$

The chemical potential of the salt, referred to its chemical potential in the standard state, is

$$\begin{aligned} (\mu_3 - \mu_3^\circ)_{\text{total}} &= (\mu_{3_e} - \mu_{3_e}^\circ) + (\mu_{3_I} - \mu_{3_I}^\circ) \\ &= (\mu_{3_e} - \mu_{3_e}^\circ) + RT \ln (N_3/N_3^\circ) = RT \ln (a_3/a_3^\circ) \\ &= RT \ln f_3 + RT \ln (N_3/N_3^\circ) \\ &= A \left(\frac{1}{D} - \frac{1}{D_0} \right) + RT \ln (N_3/N_3^\circ) \end{aligned} \quad (38)$$

Here the activity of the salt, a_3 , is defined by equation (38), and the factor A is defined in (37). The activity coefficient, f_3 , is defined by (38) and by the convention that $f_3^\circ = a_3^\circ/N_3^\circ = 1$ in a very dilute solution in the standard medium of dielectric constant D_0 . It is assumed in writing (38) that the electrostatic term is the only term that causes a deviation from

* We note that the same symbol (μ) is used to denote both chemical potential and dipole moment. It should be clear in all cases from the context which quantity is denoted.

ideal behavior. This is certainly not strictly true for any actual system, but what we are doing here is to set up equation (38) for a simple idealized model system, in order to see whether the model system resembles actual systems at all closely.

If we consider the solubility of the salt, first in the standard medium of dielectric constant D_0 , then in another medium of dielectric constant D , we have $(\mu_3)_{\text{sat}} = (\mu_3^\circ)_{\text{sat}}$ for the two saturated solutions, since both are in equilibrium with the same solid phase (see Chapter 4). Then, from (38),

$$A \left(\frac{1}{D} - \frac{1}{D_0} \right) = \left(\frac{-N\epsilon^2\nu Z_+Z_-}{2b} \right) \left(\frac{1}{D} - \frac{1}{D_0} \right) = RT \ln \frac{(N_3^\circ)_{\text{sat}}}{(N_3)_{\text{sat}}} = RT \ln f_3 \quad (39)$$

That is, the activity coefficient, f_3 , varies inversely as the ratio of the solubilities in the actual medium and the standard medium.

Consider a specific case. We take the standard state as an infinitely dilute solution of the salt in water. Let the transfer be from water ($D_0 = 80$, approximately) to ethanol ($D = 25$). Assume that both ions are univalent, and take the mean radius $b = 2 \text{ \AA} = 2 \times 10^{-8} \text{ cm}$, and $T = 300^\circ \text{ K}$. Since $\epsilon = 4.8 \times 10^{-10}$ and $k = R/N = 1.38 \times 10^{-16} \text{ erg/deg}$, we have

$$\log f_3 = \frac{23.04 \times 10^{-20}}{2.303 \times 2 \times 10^{-8} \times 1.38 \times 10^{-16} \times 300} \left(\frac{1}{25} - \frac{1}{80} \right) = 3.3$$

Thus the electrical activity coefficient, so calculated for such a hypothetical salt, is approximately 2000 times as great in alcohol as in water. From the relations previously derived (in equation 39; see also the discussion in Chapter 4) between activity coefficient and relative solubility in different media, it follows that the salt should be less soluble in ethanol than in water, by a factor of approximately 2000. Actually sodium chloride is somewhat more than 2000 times as soluble in water as in ethanol. In order to fit the experimental data for this salt by equation (39) above, we should have to take the mean ionic radius, b , as approximately $1.3 \times 10^{-8} \text{ cm}$ ($= 1.3 \text{ \AA}$), which is actually not far from the mean radius determined from X-ray studies on the sodium chloride crystal. The agreement in general is not so good as this, and our model is in many ways so crude and oversimplified that we have no right to expect any such close agreement; it is significant, however, that the calculated radii are of the right order of magnitude.

Since the addition of ethanol to water raises the activity coefficient of salt present in the medium, it follows that the addition of salt to an ethanol-water mixture must raise the activity coefficient of the ethanol.

This mutual relation between the two effects is an example of a general theorem of thermodynamics. At constant pressure and temperature, the variation of the Gibbs free energy (F) of a system of m components, expressed as a function of the masses ($n_1, n_2 \dots n_m$) of the components is

$$dF = \mu_1 dn_1 + \mu_2 dn_2 + \dots \mu_n dn_m \quad (40)$$

We shall express the masses of the components in moles. Here the chemical potential, μ_i , of the i th component is $\mu_i = (\partial F / \partial n_i)_{P, T, n_j}$ where the subscript n_j denotes that the masses of all components in the system except the i th component are held constant during the differentiation. Consider variations of chemical potentials produced by varying the masses of any two of the n components; we denote them as components i and j . We consider the effect of varying i by the amount dn_i , then the effect of varying j by the amount dn_j , starting in each case from the identical system of specified composition. Then the two variations are related by the equation

$$\frac{\partial \mu_i}{\partial n_j} = \frac{\partial^2 F}{\partial n_i \partial n_j} = \frac{\partial \mu_j}{\partial n_i} \quad (41)$$

since dF is an exact differential. This relation may also be written in terms of activities, since $d\mu_i = RT d \ln a_i$, in the form

$$\frac{\partial \ln a_i}{\partial n_j} = \frac{\partial \ln a_j}{\partial n_i} \quad (41.1)$$

Thus, if adding component j to the system increases the activity of component i , it necessarily follows that adding i to the system must increase the activity of j . The relation is reciprocal, and it holds between *any* two components in a system of m components. In our particular case the number of components is $m = 3$; we denote the water as component 1, the alcohol as component 2, the salt as component 3. Thus, we can infer immediately from (41.1) that, if addition of alcohol increases the activity of the salt (as manifested by a lowered solubility), then addition of salt increases the activity of the alcohol, as shown for instance by an increase of its vapor pressure.

If N_i is the mole fraction of component i , in a system containing m components, it may readily be shown that

$$\frac{\partial \ln N_i}{\partial n_j} = \frac{\partial \ln N_j}{\partial n_i} = \frac{-1}{\sum_{i=1}^m n_i} \quad (42)$$

Hence, since the activity coefficient of component i (f_i) may be defined

as $f_i = a_i/N_i$, it follows from (41.1) that

$$\frac{\partial \ln f_i}{\partial n_j} = \frac{\partial \ln f_j}{\partial n_i} \quad (42.1)$$

Formula (42) needs modification if some or all of the components are salts which may be considered to dissociate completely into ions. For generality, let us say that the i th component is composed of ν_i particles, so that the total number of particles is $\sum \nu_i n_i$, the sum being taken over the m components, and ν_i being unity for any nondissociating component. Then the mole fraction of component i may be written

$$N_i = \nu_i n_i / \sum_{i=1}^m \nu_i n_i \quad (42.2)$$

In this case (42) becomes

$$\frac{1}{\nu_j} \frac{\partial \ln N_i}{\partial n_j} = \frac{1}{\nu_i} \frac{\partial \ln N_j}{\partial n_i} = \frac{-1}{\sum_{i=1}^m \nu_i n_i} \quad (42.3)$$

Equation (42.1), relating the activity coefficients of the different components, is still valid in this case.

We now wish to formulate the effect of the added salt (component 3) on the activities of the water (component 1) and the other uncharged component 2 (for instance, ethanol, acetone, or ether). To do so, we must find a relation which gives the dielectric constant of a medium containing n_1 moles of component 1 and n_2 moles of component 2. For the system water-ethanol at 25° the empirical relation

$$\frac{1}{D} - \frac{1}{D_0} = \beta N_2^* = \beta n_2 / (n_1 + n_2) \quad (43)$$

where β is a constant approximately equal to 0.027, holds reasonably well over the entire range of composition. Equation (43) may be used as a first approximation for many other systems, especially over a limited range of composition when $n_2 \ll n_1$, when information is available concerning the dielectric constants of mixtures of water and component 2, the numerical value of β of course depending on the particular system under consideration. We note that N_2^* is the mole fraction of component 2 when $n_3 = 0$.

Consider now a solution containing n_1 moles of water, n_2 moles of component 2, and n_3 moles of a salt which gives ν ions in solution. By

combining equations (39) and (43) we obtain as the expression for the activity coefficient of the salt

$$RT \ln f_3 = A\beta \left(\frac{n_2}{n_1 + n_2} \right) \quad (44)$$

whence, taking account of (42.1),

$$\frac{RT \partial \ln f_3}{\partial n_1} = \frac{-A\beta n_2}{(n_1 + n_2)^2} = \frac{RT \partial \ln f_1}{\partial n_3} \quad (45)$$

Integration of the second form of this equation gives

$$RT \ln f_1 = \frac{-A\beta n_2 n_3}{(n_1 + n_2)^2} + \text{Const.} = -A\beta N_2^* \left(\frac{n_3}{n_1 + n_2} \right) + \text{Const.} \quad (46)$$

We set the constant equal to zero, which is equivalent to taking the activity coefficient of component 1, i.e., water, as unity in a solution in which the concentration of salt (component 3) is zero. It will be recalled that we have assumed that all departures of the solutions from ideality result exclusively from electrostatic effects produced by the salt ions. Then if we take account of the fact that the vapor pressure, p , is proportional to the activity, i.e., $p_1/p_1^* = f_1 N_1/N_1^*$, we may write

$$\begin{aligned} RT \ln f_1 &= RT \ln (p_1/p_1^*) + RT \ln (N_1^*/N_1) = -A\beta N_2^* \left(\frac{n_3}{n_1 + n_2} \right) \\ &= -A\beta N_2^* \left(\frac{N_3}{1 - N_3} \right) \end{aligned} \quad (47)$$

A strictly parallel procedure may be used to obtain an expression for f_2 , the activity coefficient of ethanol. The only difference is that in this case the expression for $\frac{\partial \ln f_2}{\partial n_3} = \frac{\partial \ln f_3}{\partial n_2}$ is positive instead of negative. If we adopt the same convention as before regarding standard states (this means again setting the constant of integration equal to zero) we arrive at the result

$$\begin{aligned} RT \ln f_2 &= RT \ln (p_2/p_2^*) + RT \ln (N_2^*/N_2) = A\beta N_1^* \left(\frac{n_3}{n_1 + n_2} \right) \\ &= A\beta N_1^* \left(\frac{N_3}{1 - N_3} \right) \end{aligned} \quad (48)$$

If compound 2 is only slightly soluble in water, and if water and salt dissolve little or not at all in pure component 2, then we may determine f_2 by solubility measurements at different salt concentrations. In this case, each solution is equilibrated with pure component 2, so that $p_2 = p_2^*$

in all solutions studied. Alternatively, if component 2 is very soluble in water, we may obtain an equivalent result by equilibrating the aqueous solution, with or without salt, with a nonpolar organic liquid, such as benzene, which is almost completely insoluble in water, and in which water and salt are insoluble. We then adjust the amount of component 2 in each experiment until its concentration (and therefore its activity) in the benzene phase is the same in all the solutions studied, when the

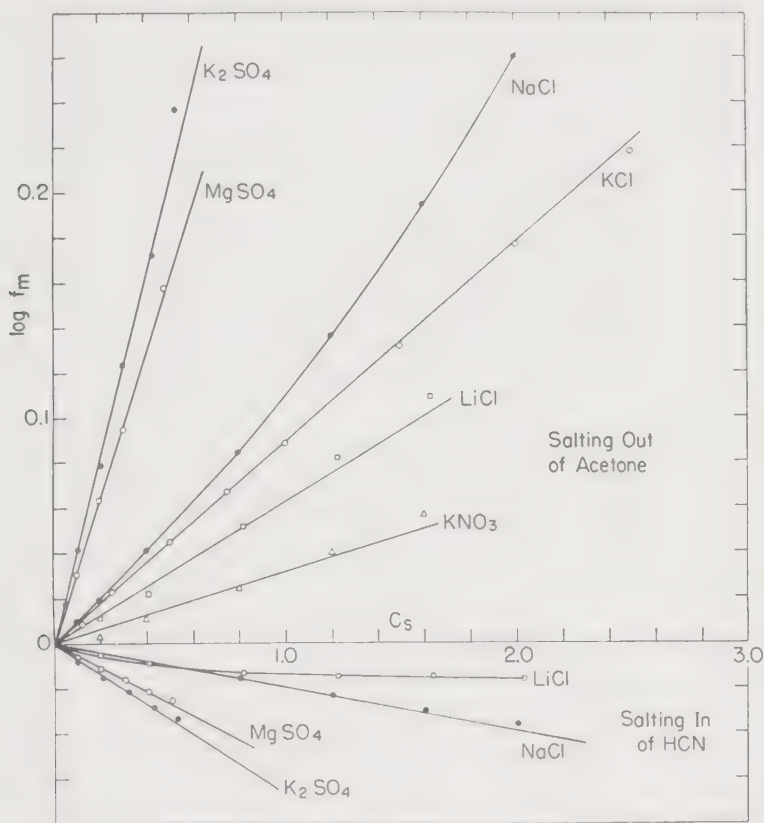


FIG. 7. The salting out of acetone, which lowers the dielectric constant of water, and the salting in of HCN, which raises the dielectric constant. [From the work of Gross and Schwartz (1930), and Gross and Iser (1930).]

benzene phase and the aqueous phase are in equilibrium. Then the activity (vapor pressure) of component 2 is a constant in all the different aqueous salt solutions, and the change in its solubility, as salt is added, reflects the change in activity coefficient. This is the method used by Gross (1930) in studies on acetone and on hydrocyanic acid in water (Fig. 7). Then from (48) we have for the change in solubility of component 2, expressed as mole fraction, with increase of n_3 ,

$$\frac{\partial \ln (N_2/N_2^*)}{\partial n_3} = - \frac{\partial \ln f_2}{\partial n_3} = -A\beta N_1^*/RT(n_1 + n_2) \quad (49)$$

We note that $\partial \ln f_2 / \partial n_3$, as given by (49), is equal to $\partial \ln f_3 / \partial n_2$; the latter coefficient can be obtained by differentiating the expression for $\ln f_3$ in (38) with respect to n_2 , making use of (43). Thus the reciprocal relation between the activity coefficients, required by (42.1), is fulfilled. It is easy to verify the same relation with respect to components 1 and 3.

If n_1 is very much greater than n_2 or n_3 , then $N_1^* \cong 1$, and $n_1 + n_2 \cong n_1$. If we choose n_1 as the number of moles in 1 kg of component 1 (water), then n_3 becomes the molality of the salt, m_3 . Moreover the *ratio* of two solubilities of component 2, N_2/N_2^* , in two different media is virtually identical with the *ratio* of the two solubilities, S_2/S_2^* , expressed as moles per liter or moles per kilogram of water, if n_2 and n_3 are both very small

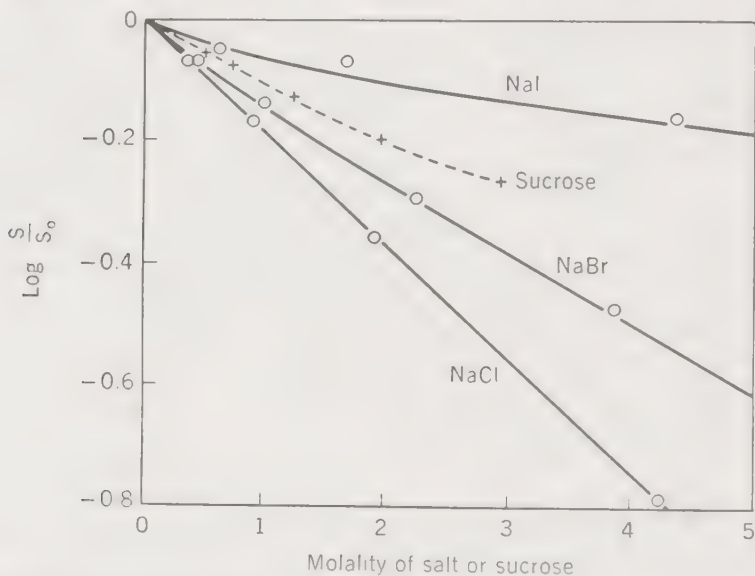


FIG. 8. Salting out of ethyl acetate. (From the data of Glasstone and Pound, 1925, as represented by Scatchard, 1927.)

compared to n_1 . With these approximations we can write, converting to denary logarithms,

$$\partial \log (S_2/S_2^*)/\partial m_3 = -A\beta/2.303RT = -K_s \quad (50)$$

Here K_s is called a salting-out constant. This relation has been derived as a limiting equation, valid only at very low salt concentrations when the electrical free energy of the system is a linear function of n_3 . The linear relation obtained by integrating equation (50), however, on the assumption that K_s is a constant, is often found empirically to hold over much wider ranges of salt concentrations than we should have any right to expect; sometimes up to 2 or 3 *M* salt. Examples of some data are shown in Figs. 7 and 8, taken from the work of Gross (1930) on acetone and hydrocyanic acid in aqueous salt solutions, and of Glasstone and

Pound (1925) on ethyl acetate in similar solutions. As an empirical relation to describe the data of such experiments, we write therefore

$$\log (S_2/S_2^*) = -K_s m_3 \quad (51)$$

Alternatively, the salting-out equation may be written in terms of the ionic strength of the solution, ω , instead of m_3 . The ionic strength is defined⁶ by taking the molar concentration (or molality) of each ion and multiplying by the square of its valence (Z_i^2):

$$\omega = \frac{1}{2} \sum C_i Z_i^2 \quad \text{or} \quad \omega = \frac{1}{2} \sum m_i Z_i^2 \quad (52)$$

If a salt dissociates to give ν_+ ions of valence Z_+ , and ν_- ions of valence Z_- , then if m is the total molality of salt, the corresponding ionic strength is

$$\omega = \frac{m}{2} (\nu_+ Z_+^2 + \nu_- Z_-^2) = -\frac{m}{2} (\nu_+ Z_+ Z_-) \quad (53)$$

where $\nu = \nu_+ + \nu_-$. Compare the discussion of equation (37).

Thus the salting-out effect may also be written from (51), (52), and (53), in terms of the ionic strength:

$$\log (S_2/S_2^*) = -K_s' \omega \quad (54)$$

where $K_s' = 2K_s/(-\nu Z_+ Z_-)$, if all the ionic strength is due to a single salt. If the salt is uni-univalent, like NaCl, $m_3 = \omega$ and $K_s' = K_s$. If the salt is uni-bivalent, like CaCl_2 or Na_2SO_4 , $\omega = 3m_3$ and $K_s' = K_s/3$. If it is bi-valent, like MgSO_4 , $\omega = 4m_3$ and $K_s' = K_s/4$.

It may be seen from Figs. 7 and 8 that the agreement of the experimental data with the linear relations (51) and (54) is not by any means exact. The equations are empirically useful, however, and the mean radii for the salt ions calculated from them are of the right order of magnitude. Data on salting-out constants are given in Table I. We note that β in equation (50) and earlier equations is given experimentally by dielectric constant measurements on mixtures of components 1 and 2, whereas A is inversely proportional to the mean ionic radius and otherwise depends only on universal constants and on the valence type of the salt. If β is negative, as with hydrocyanic acid in water, "salting in" may be obtained instead of salting out. The exact values of the calculated ionic radii should not be taken too seriously, for somewhat different values are ob-

⁶ It has been common to denote the ionic strength, written in terms of molality, by the symbol μ , which we avoid here because we have used μ to denote dipole moment. The ionic strength in moles per liter is often denoted by $I/2$, but we have chosen a briefer symbol, ω . We shall use the same symbol to denote ionic strength either in terms of moles per liter or in terms of molality. At low ionic strengths, in aqueous solutions, the two units become nearly identical. Under other conditions we specify in each case, as it arises, the units of concentration to be employed.

TABLE I
VALUES OF SOME SALTING-OUT CONSTANTS, K_s' , FOR GASES,
NONELECTROLYTES, AMINO ACIDS, AND PROTEINS
(Temperature 25°)

Substance	NaCl	MgSO ₄	(NH ₄) ₂ SO ₄	Na ₂ SO ₄	Phos- phate
Oxygen	0.13				
Hydrogen sulfide	0.06				
Acetylene	0.10	0.06		0.10	
Phenylthiourea	0.16	0.07		0.12	
Ethyl acetate	0.18				
Cystine			0.05		
α-Aminobutyric acid	0.04				
Leucine	0.09				
Tyrosine*	0.31				
β-Lactoglobulin				0.63	
Hemoglobin (horse)		0.33	0.71	0.76	1.00
Hemoglobin (man)					2.00
Myoglobin			0.94		
Egg albumin			1.22		
Fibrinogen	1.07		1.46		2.16

Values for the first four substances listed are from M. Randall and C. F. Failey, (1927). Value for ethyl acetate from S. Glasstone and A. Pound (1925). Other values from E. J. Cohn and J. T. Edsall (1943).

* The value for tyrosine given here is probably considerably too high. New and careful measurements by D. B. Wetlaufer, *Compt. rend. trav. lab. Carlsberg Sér. chim.* **30**, 135, 1956 give $K_s' = 0.062$ for tyrosine in sodium chloride.

tained by a more refined theory, due to Debye (1927), which is discussed below; and even this more refined theory is still only an approximation to the actual complex physical situation. The facts, however, justify the general conclusion that the forces which underlie the salting-out effect are primarily electrostatic.

We may also formulate the effect of adding salt on the vapor pressures (or, more generally, on the activities) of components 1 and 2, in a one-phase system of components 1 and 2, to which salt is added. We divide (47) and (48) by RT , and differentiate these two equations with respect to n_3 , making use of (42) and (42.3):

$$\begin{aligned} \frac{\partial \ln (p_1/p_1^*)}{\partial n_3} &= \frac{-A\beta}{RT} \frac{N_2^*}{(n_1 + n_2)} + \frac{\partial \ln N_2}{\partial n_3} \\ &= \frac{-A\beta}{RT} \frac{N_2^*}{(n_1 + n_2)} - \frac{\nu}{\sum \nu_i n_i} \end{aligned} \quad (55)$$

$$\frac{\partial \ln (p_2/p_2^*)}{\partial n_3} = \frac{A\beta}{RT} \frac{N_1^*}{(n_1 + n_2)} - \frac{\nu}{\sum \nu_i n_i} \quad (56)$$

The salt addition always lowers the vapor pressure of component 1—except for the unusual cases in which β is negative—for the addition of salt decreases both the activity coefficient and the mole fraction of component 1. For component 2, however, the activity coefficient is increased but the mole fraction is lowered by the addition of salt. If component 2 is of low polarity, so that β is large, and if the salt is of high valence type, with small ionic radii, so that A is large, the first term on the right of (56) predominates, and the vapor pressure of component 2 is then markedly increased by the addition of salt, especially if N_1^* is not much below unity. Indeed most molecules of low polarity—alcohols, ethers, and ketones, for instance—show marked positive deviations from Raoult's law in water, so that p_2^* may be a large fraction of p_2° , the vapor pressure of the pure organic liquid, even when N_2^* is rather small (see Chapter 4, p. 188). Often in such cases the salt addition leads to a considerable increase in the total vapor pressure of the system, the increase in p_2 being much greater than the decrease in p_1 . In a two-component system, of course, the addition of a salt or other nonvolatile solute always leads to a decrease in the vapor pressure of the solvent. The fact that such a vapor pressure increase can and does occur in a three-component system of the sort we have just been discussing has in the past appeared paradoxical to many chemists, but it is a simple consequence of the salting-out effect.

SALTING OUT OF PROTEINS

At high salt concentrations, as we have already indicated, proteins are salted out by many neutral salts, and over a wide range of solubility $\log S$ is a linear function of the ionic strength. The validity of this relation was pointed out by E. J. Cohn (1925); the accuracy with which it holds for many proteins is illustrated by some typical data shown in Fig. 9. The use of salting out as a method of separating proteins from the complex mixtures in which they occur in nature has been known as an empirical procedure for over a century; it still remains one of the most important of all the techniques of protein fractionation. Concentrated solutions of ammonium sulfate, sodium sulfate, or phosphate buffers have generally proved the most useful substances for this purpose, because they have high values of the salting-out constant, K_s . It may be readily seen from Fig. 9 how sharply separated the regions of precipitation often are for different proteins; the values for fibrinogen and serum albumin, both found in blood plasma, may be noted as an extreme example.

The salting out of proteins differs from that of small uncharged molecules, since it is generally observed only at high salt concentrations. When small additions of salt are made to a solution of a protein dissolved in pure water, its activity coefficient is generally diminished; that is, its

solubility increases. This phenomenon, often known as "salting in," is due to the interionic forces, or the forces of attraction between ions and dipolar ions, which are discussed later in this chapter. It is only at high values of m_3 that an equation of the form of (51) or (54) holds for proteins; in this case, therefore, S_2^* is not an actual solubility, but a hypothetical extrapolated value obtained by extending the observed linear

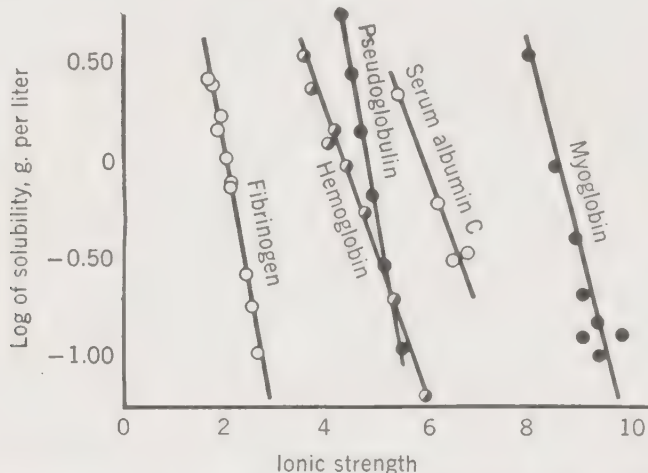


FIG. 9. The salting out of several proteins, illustrating the logarithmic relation defined by equation (57). (From Cohn and Edsall, 1943.)

portion of the curve back to $m_3 = 0$. To emphasize this difference, the equation may be written in a somewhat different form:

$$\text{Log } S = \text{Log } S_i - K_s m_3 \quad \text{or} \quad \text{Log } S = \text{Log } S_i - K_s' \omega \quad (57)$$

where S_i is the hypothetical ideal solubility at $m_3 = 0$. For a given protein and a given salt, K_s and K_s' are found experimentally to be independent of temperature and also of pH , in the cases so far studied; however, S_i is a function of both these variables. These statements are illustrated by the data of Fig. 10, taken from the work of A. A. Green on horse carboxyhemoglobin. The slopes of all the lines are the same within experimental error, but the lines are displaced up or down by variations of pH or temperature. The solubility is a minimum at or near the point at which the net electric charge on the protein molecule is zero, and rises as the charge increases in either the positive or negative direction. Rise of temperature *decreases* the solubility of hemoglobin in these concentrated salt solutions, although the effect of temperature is in the opposite direction for most proteins dissolved in dilute salt solutions; these temperature variations are related to the heat of solution of the solid protein in the medium (Chapter 4). The fact that K_s is independent of electric charge

on the molecule and of temperature indicates that it depends on characteristics of the molecule such as its size and shape, rather than on its electrical properties. Indeed K_s is commonly largest for large molecules such as fibrinogen (molecular weight near 340,000), and smaller for smaller molecules such as myoglobin (molecular weight near 17,000); but this is not an invariable rule. A list of salting-out constants for various salts and various molecules which are salted out, both small and large, is given in Table I. It will be seen that the relative values of K_s , for different salts, are in the same order for any given molecule, although the absolute values differ greatly according to the size and nature of the molecule which is being salted out.

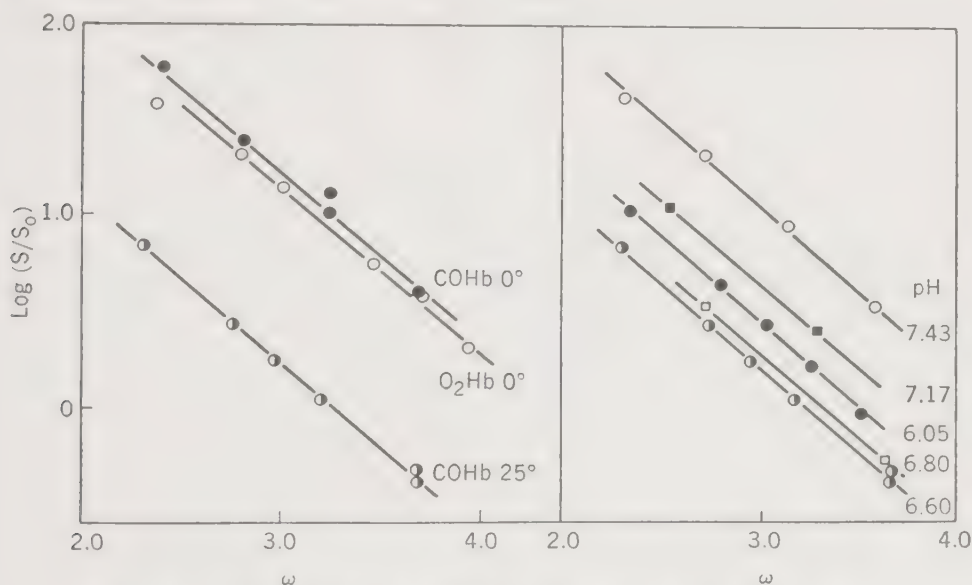


FIG. 10. The salting out of hemoglobin. (From A. A. Green, 1931.)

It was noted in the late nineteenth century by the distinguished German biochemist Franz Hofmeister that different salts could be placed in a regular order with respect to their salting-out effect on proteins, the order remaining essentially the same for different proteins. He pointed out that the effect of anions and cations of different salts could be regarded as essentially additive; and the order in which he listed the ions is still known as the Hofmeister series. It was at one time thought that the phenomena observed by Hofmeister had to do in some particular way with the properties of colloids. Since the same series is found, however, for the action of salts on simple compounds such as ethyl acetate, it is clear that the significance of the series is quite unrelated to colloidal chemistry in any special sense. It is rather a general function of the size and hydration of ions.

EFFECTS OF DIPOLAR IONS

Salting-out effects can be produced by dipolar ions as well as by ions. Perhaps the term "salting-out" is technically a misnomer in describing such systems, since dipolar ions are not salts. Since a dipolar ion contains at least one anionic and one cationic group, however, bound together but separated by a considerable distance, the interactions of these ionic groups with other nonpolar or moderately polar organic molecules are qualitatively very similar to those of salts. It is a reflection of this fact that the aliphatic amino acids, like salts, are always much less soluble in organic solvents than in water. Of course it follows from this that the addition of an amino acid to a mixture of water (component 1) and an organic solvent (component 2) causes a decrease in the solubility, or an increase in the vapor pressure, of component 2. This is illustrated by the work of England and Cohn (1935) on the distribution coefficients of amino acids between water and *n*-butanol, in which they also determined the converse effect—namely, that of the added amino acid on the solubility of *n*-butanol in water. These three-component systems form two phases—a lighter phase containing much butanol, less water, and very little of the amino acid, and a denser phase containing chiefly water and amino acid. The solubility of butanol in the lower phase, when no amino acid is present, is approximately 1 *M* at 25°. The addition of glycine to the water–butanol mixture markedly decreases the amount of butanol in this phase; the butanol solubility drops to approximately 0.65 *M* in 0.7 *M* glycine, and to less than 0.50 *M* in 1.5 *M* glycine. The effects of alanine and of α -aminobutyric acid on the solubility of butanol in water are in the same direction but less marked than the effect of glycine. Qualitatively the effect of adding an amino acid such as glycine to the butanol–water system is just like that of adding a salt. The solubility of dipolar ions in water and organic solvents is discussed in its broader aspects later in this chapter.

DEBYE'S THEORY OF SALTING OUT BY REDISTRIBUTION OF SOLVENT MOLECULES AROUND THE IONS

An alternative theoretical treatment of the salting-out effect, which is in many respects closer to molecular reality, was given by Debye (1927). This is based on the fact that in a mixture of water (1) and an organic solvent (2) the introduction of ions causes a rearrangement of the molecules in the mixed solvent, the water molecules crowding around the ions and squeezing out the molecules of component 2. Quantitatively this may be formulated in terms of a general equation which gives the energy stored in a dielectric medium which is acted on by an electric

field. It may be readily shown⁷ that this energy is equal to $DE^2/8\pi$ per unit volume, where E , as usual, is the electric intensity. In a medium which is very dilute in ions, the value of E at any point is simply given by the relation $E = Z\epsilon/Dr^2$, $Z\epsilon$ being the charge on the nearest ion to the point in question and r the distance from the point to the center of the ion.⁸ Thus we may write, for the electrostatic energy stored in a volume element dV ,

$$(\text{Energy})_e = \frac{Z^2\epsilon^2}{8\pi Dr^4} dV \quad (57.1)$$

Let the system before addition of salt contain n_1 moles of component 1 and n_2 moles of component 2 per milliliter, the partial molal volumes of these components being \bar{V}_1 and \bar{V}_2 , respectively. Since the two kinds of molecules between them fill the volume, the relation must hold (see Chapter 4, equation 39) that

$$n_1\bar{V}_1 + n_2\bar{V}_2 = 1 \quad (57.2)$$

The mole fractions $N_1^* = n_1/(n_1 + n_2)$ and $N_2^* = n_2/(n_1 + n_2)$ are defined as before. The introduction of salt ions leads to a redistribution of components 1 and 2 in regions where the field strength is high, since from (57.1) the electrostatic energy is diminished, for given values of $Z\epsilon$ and r , by increasing the value of D . If the dielectric constant of pure component 1 is higher than that of 2, this means that component 1 will tend to crowd around the ions, thereby increasing D in regions where E is high. On the other hand, when the concentration (or mole fraction) of a component is different in a particular region from its average value in the solution as a whole, work is involved in producing this change in concentration. The associated change in free energy, if n_i moles of component i are involved, is $n_iRT \ln (N_i/N_i^*)$ if we consider the solution as behaving in this respect like an ideal solution.

The introduction of salt into the system which originally consisted only of components 1 and 2 thus changes the free energy of the system by setting up an electrostatic field, and by the associated concentration changes of components 1 and 2 in the neighborhood of the ions. In a

⁷ We shall not attempt to prove this proposition here, since it involves certain mathematical concepts which we have not introduced in this chapter. The proof, however, is given in practically all texts on electricity and magnetism—see general references given at the end of this chapter, such as Page and Jeans.

⁸ This relation is valid because the ions are very far apart and only the field due to the nearest neighboring ion at any point need be considered. If a given point is equally distant from two or more ions, we may assume that the value of E at such a point is so small, because the ions are so far away, that the small region in question makes a negligible contribution to the total energy stored in the medium.

volume element dV , at a distance r from the nearest ion of charge $Z\epsilon$, and containing n_1 moles of component 1 and n_2 moles of component 2, the total free energy change is

$$(F - F^*) dV = \left[n_1 RT \ln (N_1/N_1^*) + n_2 RT \ln (N_2/N_2^*) + \frac{Z^2 \epsilon^2}{8\pi D r^4} \right] dV \quad (57.3)$$

The total free energy change in the whole volume of solution, due to the addition of the ions of the salt, is

$$\Delta(F - F^*) = \int (F - F^*) dV \quad (57.4)$$

the integral being taken over the total volume of the system. We know from the discussion of free energy in Chapter 4 that, in a system at equilibrium, at constant pressure $\Delta(F - F^*)$ will assume a minimum value, subject to the condition that the volume is filled up by the molecules, and therefore that (57.2) must hold,⁹ and to the further conditions that the total number of molecules of components 1 and 2 in the whole system remains constant—that is, that $\int n_1 dV = \text{a constant}$, and $\int n_2 dV = \text{another constant}$, when integration over the whole volume is carried out.

These considerations lead to the equation for the distribution of component 2 around any given ion taken as center:

$$N_2/N_2^* = C_2/C_2^* = e^{-R^4/r^4} \quad (57.5)$$

Here C_2 is the concentration of component 2 in moles per liter; the relation $C_2/C_2^* = N_2/N_2^*$ holds if $N_2^* \ll 1$, which we assume here. The term R is a characteristic distance which is given, if N_2^* is small and the salt concentration low, by the equation

$$R^4 = \frac{1000\epsilon^2 Z^2}{8\pi kT} \left(\frac{1}{D} - \frac{1}{D_0} \right) \frac{1}{C_2}$$

⁹ Two assumptions are involved here: (1) that \bar{V}_1 and \bar{V}_2 are independent of N_1^* and N_2^* , and (2) that no change in the volume occupied by components 1 and 2, in the system as a whole, takes place as a result of the introduction of the ions of the salt. Both assumptions are in general incorrect. Data showing the incorrectness of the first assumption, for alcohol-water mixtures, are cited in Chapter 4; most solutions of organic molecules in water show comparable variations in \bar{V}_1 and \bar{V}_2 with composition. The second assumption is also generally incorrect, since the intense electrostatic fields around the ions lead to closer packing of the surrounding water molecules and other dipoles into the available space. (See the discussion of water in Chapter 2, and the discussion of electrostriction in Chapter 4, p. 173.) The errors involved in considering (57.2) to be generally valid, however, are probably small compared to those involved in other approximations made in this treatment.

For example, dioxane–water mixtures obey the approximate relation $D_0/D = 1 + 0.091C_2$ at 1.4° (Scatchard and Benedict, 1936), and the corresponding value of R for dioxane in water is 2.53 Å. In such a case the concentration of a substance such as dioxane in water diminishes very rapidly as the surface of the ion is approached. Even in mixtures containing only a very small amount of water, the shell of solvent immediately around the ions will be made up almost entirely of water molecules.

We shall not go into further details here concerning this theory of salting out. The general molecular picture which it gives of the distribution of the two kinds of solvent molecules around the ion is certainly much closer to the actual molecular situation than is the simple picture, that we have employed earlier, of the solvent as a continuous structureless medium of uniform dielectric constant D . Moreover, the Debye theory has been found by several workers—see for instance Gross (1930) and Scatchard and Benedict (1936)—to give more reasonable results than the earlier theory for the molecular radii of the ions. This theory, too, has its difficulties, however. Electrostriction effects are neglected, and the dielectric constant of the solvent is regarded simply as a function of the composition, without allowing for effects due to the reorientation of solvent molecules around the ions. Moreover, the predicted change in the concentration of compound 2 with increasing distance from the center of the ion (by equation 57.5) is so rapid that the theory predicts a large concentration change within a distance of the order of a single molecular diameter. Obviously this is physically impossible, unless we regard this theory of molecular distribution in purely statistical terms. A critical discussion is given by Scatchard (1941).

Another way of interpreting the salting-out effect is in terms of the image charges produced by the ions in the surrounding neutral molecules, which may be regarded as if they were cavities of low dielectric constant, immersed in the solvent medium. These charges, as we have seen, give rise to a dipole, with its positive pole pointing toward a positive ion; hence the ion and the dipole repel one another. If the cavity were of higher dielectric constant than the surrounding medium, then the orientation of the image dipole, and consequently the whole effect, would be reversed. This relation between image charges and the salting-out effect will appear later, in Kirkwood's theory of the interactions of ions and dipolar ions.

It may be seen from this discussion that there is still need for a more adequate theory of the salting-out effect, preferably framed in terms of the dimensions, dipole moments, and electrical polarizabilities of the individual molecules, rather than in terms of the macroscopic dielectric constant of the whole medium. There is also great need for more experi-

mental data; systematic measurements of K_s in relation to structure, for homologous series of alcohols, ethers, esters, ketones, and other classes of organic molecules in water, scarcely exist at present. They would be of great value in providing empirical relations between K_s and structure, for various organic molecules and various salts. They would also provide, in conjunction with suitable measurements of dielectric constants of the solutions involved, essential material for the further development of the theory. Even at present, however, the salting-out effect is recognizable as a phenomenon of major importance in chemistry and biochemistry, and a striking illustration of the importance of electrostatic interactions.^{9a}

Ionic Interactions and the Debye-Hückel Theory

In discussing the salting-out effect, we pictured the ions as charged spheres, immersed in a continuous medium of dielectric constant D , neglecting interactions between the ions. The same general concepts may be extended to systems in which the ions are close enough to interact significantly. We propose now to calculate the effect of varying the total concentration of the ions on the activity coefficient of one particular electrolyte component of the medium. If we take the activity coefficient of the electrolyte as unity at infinite dilution in the pure solvent medium—usually water—then the activity coefficient of the electrolyte decreases with increasing electrolyte concentration. This is always true in the range of moderate concentrations—say below 0.5 M , or thereabouts. The same is true for systems containing several different kinds of ions, as is shown most directly by solubility measurements. Slightly soluble salts, such as thallous chloride (Tl^+Cl^-) or barium iodate ($Ba^{++}IO_3^{--}$) become more soluble when other salts without a common ion are added to the system. The effect of a given added salt on the solubility of barium iodate is much greater than its effect on the solubility of thallous chloride; in general, the solubility of salts with ions of higher valences is increased much more than that of salts with univalent ions. We know already (see Chapter 4) that increase of solubility means a decrease of activity coefficient; the same decrease can be shown also by other methods, such as freezing point determinations and electromotive force measurements.

Qualitatively, the reason for these effects is not far to seek in terms of the fundamental electrostatic concepts already given, and indeed it has already been hinted at. It will be recalled that in calculating the effect of the dielectric constant on the activity of an ion we treated the ion as a conducting sphere and calculated its electrical energy from the self potential of the sphere at its surface (see p. 253), pointing out that we were neglecting the effect of the presence of surrounding ions. If the ionic con-

^{9a} For more extensive data and critical discussion, see also Long and McDevit (1952).

centration of the medium is appreciable, we must expect the presence of surrounding ions to make itself felt. There must, on the average, be a tendency for a given ion to be surrounded in its immediate vicinity by an excess of ions of opposite charge. These have a depressing effect on the electrical potential at the surface of the given ion, and there results a diminution in its electrical energy. This, of course, means that there is a decrease in the chemical potential of the ion and, consequently, in its activity coefficient, in the presence of other ions.

In order to formulate this quantitatively we consider the work of transferring the ions of a molecule of a given strong electrolyte from an infinitely dilute solution in the pure solvent, in which its activity coefficient, as well as that of the ions, may be taken as unity, to a more or less concentrated solution in the same solvent, which may contain other electrolytes as well. Conceptually the transfer process may be broken down into three steps, each carried out reversibly. First, we discharge the ions. Second, we transport them in their uncharged condition to the more concentrated solution. Third, we recharge them in their new environment. The first step involves a negative amount of work equal to the electrical energy of the ions under conditions where the effect of surrounding ions may be neglected. Under such conditions the electrical energy of each ion of radius b is one-half the product of its charge ($Z\epsilon$) times its self potential ($Z\epsilon/Db$) as was shown in the earlier calculations relating to the salting-out effect (p. 265). The second step involves a positive amount of nonelectrical work depending only on the initial and final concentrations of the ions. The third step involves the electrical work of recharging the ions, this time, however, in a solution of sufficient concentration so that the effect of neighboring ions is appreciable. This work of recharging may itself be expressed as the sum of two terms, one equal to the electrical energy of the ions in the absence of their neighbors, and one equal to the additional energy due specifically to the presence of the neighbors. The former term exactly cancels the electrical work of the first step, and the latter then remains as the only net *electrical* work term of the whole process. It is this which gives rise to the effect in which we are interested, that is, the reduction in the activity coefficient of an electrolyte in a solution containing a finite concentration of electrolytes. If we denote this additional work term by W , then it follows from fundamental definitions that the activity coefficient of the electrolyte in question, *due to its presence in an ionic environment*, is given by

$$W = RT \ln f \quad (58)$$

There may, of course, also be a change of activity due to nonelectrical effects, but this need not concern us here.

Now, W , it should be insisted, is only that portion of the electrical

work of recharging all the ions of a molecule of the strong electrolyte under consideration which results from the presence of the surrounding ions. The contribution to W due to recharging any one of the ions of the electrolyte is the work involved in gradually increasing the charge on the ion by small increments, de , which is given by the integral

$$W = \int_0^{Z\epsilon} \psi_i de \quad (59)$$

Here $Z\epsilon$ is the full charge of the ion and ψ_i is the potential at its surface, at a given stage of the charging process, resulting from the surrounding ions. This expression is the same as that used in earlier calculations except that here ψ_i takes the place of the self potential of the ion. Like the self potential, ψ_i will be a function of the charge of the ion at any stage of the charging process, and, as we shall see, it varies from zero to a final value as the process proceeds and an ion atmosphere builds up about the ion in question. Since the ion atmosphere is predominantly made up of ions of opposite charge to that of the ion in question, ψ_i is of opposite sign to the self potential of the ion.

In order to determine ψ_i , the potential at the surface of an ion due to the presence of the surrounding ion atmosphere, it is necessary to solve the far more general problem of determining the total potential due to the ion *and* its ion atmosphere at *any* point outside the ion. To deal with this problem we treat the ions as conducting spheres and assume that the distance of closest approach of any two ions, measured between centers, is a , the same for any pair of ions. The distance a is necessarily a kind of average quantity. For distances greater than a , we assume that the forces between ions are purely electrostatic, so that the distribution of the ions is determined solely by their electrostatic energy. We fix our attention on a given ion, k , of radius b_k , and introduce a system of spherical coordinates with its origin in the center of this ion (Fig. 11). On the average, the neighboring ions will be distributed about the ion k with spherical symmetry, and the potential must, therefore, be spherically symmetrical about the origin. We treat the charge of each surrounding ion as if it were a point charge located at the center of that particular ion. In the region between $r = b_k$ and $r = a$, we assume that there is no charge, since the center of no ion can penetrate there. Outside the region ($r \geq a$) we treat the medium as continuous and as the seat of a space charge determined at any point by the concentrations of the various ions. Everywhere for $r > b_k$ we apply the laws of macroscopic electrostatics, assuming the dielectric constant to be constant and equal to the macroscopic value of the pure solvent. This is our model, somewhat crude, to be sure, but one which leads to results that accord with experience and give it a new significance.

We denote the total potential, at any point in the region around the ion, by ψ . It may be expressed as

$$\psi = \psi_k + \psi_i \quad (60)$$

where ψ_k is the potential due directly to the central ion k under consideration, and ψ_i is the potential due to the ion atmosphere (this is not, of course, restricted to its value at the surface of the ion k as in equation

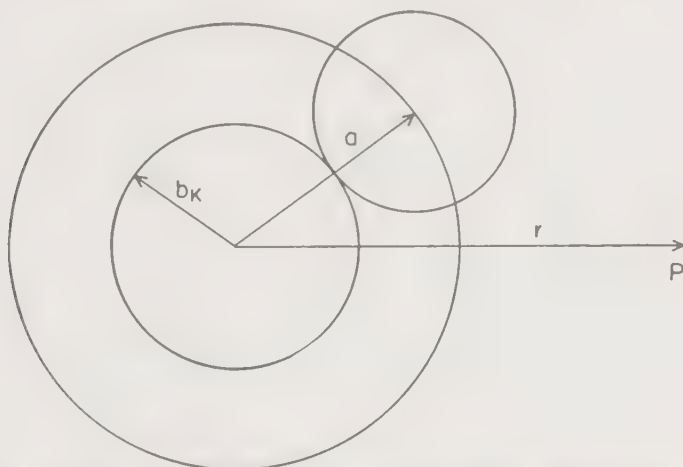


FIG. 11. The central ion of radius b_k and the collision diameter a which represents the distance of closest approach between the centers of two ions.

59). Outside the surface $r = a$, ψ will be a solution of Poisson's equation (14), which has already been discussed:

$$\nabla^2 \psi = \frac{-4\pi\rho}{D}$$

Here ρ denotes the density of the space charge resulting from the unequal concentrations of positive and negative charge in the ion atmosphere. On the other hand, in the region inside this surface but outside the ion ($b_k \leq r \leq a$), ψ must be a solution of Laplace's equation:

$$\nabla^2 \psi = 0$$

for in this region the charge density is everywhere zero. As we have said, the average distribution of surrounding ions and therefore the potential, ψ , must be spherically symmetrical about the origin.

Consider first the inner region ($b_k \leq r \leq a$) where Laplace's equation holds. Here a solution may be obtained very simply from the expression for the electric intensity. In order to determine the electric intensity we construct a spherical surface of radius r which may have any value between b_k and a . Because of spherical symmetry the electrical intensity,

E , must be constant over this surface, and, therefore, by Gauss's law (equation 8.1) it is given by

$$E = \frac{Z_k \epsilon}{Dr^2} = - \frac{d\psi}{dr}$$

for the only charge enclosed by the surface is the charge of the central ion, k . It follows that in this region ψ is given by

$$\psi = \frac{Z_k \epsilon}{Dr} + A \quad (\text{if } b_k \leq r \leq a) \quad (61)$$

where A is a constant.

Let us turn now to the outer region ($r > a$) where Poisson's equation holds. Because of spherical symmetry this equation assumes the form already given in equation (13):

$$\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{d\psi}{dr} \right) = - \frac{4\pi\rho}{D}$$

No such simple device as that employed in the preceding paragraph is applicable here. In order to integrate equation (13), we must find an expression for ρ in terms of r and ψ . Such an expression can be obtained with the aid of the Boltzmann distribution law. Suppose that there are altogether i species of ions present and that the valence of an ion of the i th species is Z_i , positive or negative. Suppose also that the average number of ions of the i th species in the whole solution, per unit volume, is n_i . Then the electrical energy of an ion of species i at any point where the potential is ψ is given by $Z_i \epsilon \psi$. By Boltzmann's law, the number of such ions in an element of volume dV where the potential is ψ is

$$dn_i = n_i e^{-Z_i \epsilon \psi / kT} dV$$

It follows that the mean density of charge is

$$\rho = \sum_i Z_i \epsilon n_i e^{-Z_i \epsilon \psi / kT}$$

When this value of ρ is substituted in Poisson's equation, the result cannot be integrated. Consequently, we replace it by an approximate expression obtained by expanding the exponentials in series and retaining only the first two terms:

$$\rho \cong \sum_i Z_i \epsilon n_i \left(1 - \frac{\epsilon Z_i \psi}{kT} \right) \quad (62)$$

This is justifiable for small values of the argument but limits the applicability of the results to systems in which $Z_i \epsilon \psi$ is small compared to kT .

Since the solution as a whole is electrically neutral, $\sum Z_i \epsilon n_i = 0$ ¹⁰ and (62) becomes

$$\rho \cong \sum \frac{-n_i Z_i^2 \epsilon^2}{kT} \psi \quad (63)$$

When this is introduced into Poisson's equation, the result is

$$\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{d\psi}{dr} \right) = \kappa^2 \psi \quad (64)$$

where κ , which depends only on the composition of the solution, is given by

$$\kappa^2 = \frac{4\pi\epsilon^2}{DkT} \sum n_i Z_i^2 \quad (65)$$

Equation (64), valid everywhere for $r > a$, may be integrated by straightforward methods. Since it involves the second derivative of ψ with respect to r , two integrations are involved, and the general solution must contain two arbitrary constants. The result, as may be readily verified by differentiation, is

$$\psi = \frac{B e^{-\kappa r}}{r} + \frac{C e^{\kappa r}}{r}$$

where B and C are constants. Since the potential associated with any ion must approach zero as the distance from the ion increases indefinitely, it is clear that the constant C must be set equal to zero, so that we are left with

$$\psi = \frac{B e^{-\kappa r}}{r} \quad (\text{for } r \geq a) \quad (66)$$

In order to complete the problem it is necessary to determine the two constants A (equation 61) and B (equation 66). This may be achieved by taking account of the boundary conditions which prevail at the surface of discontinuity at $r = a$, where the density of the space charge suddenly rises from zero to a finite, if small, value. Although there is a discontinuity in the value of ρ at $r = a$, we have already shown (see the

¹⁰ This is not *strictly* true, for we are considering the solution surrounding the central ion, of charge $Z_k \epsilon$, and if the solution as a whole is to be neutral, the net charge in the surrounding medium must be numerically equal, and opposite in sign to $Z_k \epsilon$. Compared with the total number of charges in the solution, however, $Z_k \epsilon$ is generally so small that the error in the assumption that $\sum Z_i \epsilon n_i = 0$ is negligible. For a central ion of very large net charge, however, such as a highly charged protein or polyelectrolyte ion, in a medium of very low salt concentration, the error in the assumption would be appreciable, and a somewhat different procedure would have to be adopted.

discussion of equation 13.2) that the electric intensity, E , and the potential, ψ , are continuous functions of r throughout this region. Consequently, at the surface $r = a$, we can equate the values of ψ and $d\psi/dr$ given by equations (61) and (66). This yields the following two equations as the boundary conditions involving ψ and $d\psi/dr$, respectively:

$$\frac{Z_k \epsilon}{Da} + A = \frac{Be^{-\kappa a}}{a}$$

and

$$B = \frac{Z_k \epsilon e^{\kappa a}}{D(1 + \kappa a)} \quad (67)$$

The constant B is thus given directly, and A follows at once:

$$A = \frac{-Z_k \epsilon}{D} \frac{\kappa}{1 + \kappa a} \quad (68)$$

The problem of calculating the total potential is now complete. In the outer region ($r \geq a$) the introduction of B into (66) gives

$$\psi = \frac{\epsilon Z_k}{Dr} \left(\frac{e^{-\kappa(r-a)}}{1 + \kappa a} \right) \quad (\text{for } r \geq a) \quad (69)$$

This may be compared with the expression

$$\psi = \frac{\epsilon Z_k}{Dr}$$

which would hold in the absence of the ion atmosphere. The actual potential is always less than the latter, into which it degenerates as $\kappa \rightarrow 0$, corresponding to the case of an infinitely dilute solution in the pure solvent.

In the inner region $b_k \leq r \leq a$ the introduction of A from (68) into (61) gives

$$\psi = \frac{Z_k \epsilon}{Dr} - \frac{Z_k \epsilon}{D} \left(\frac{\kappa}{1 + \kappa a} \right) \quad (\text{if } b_k \leq r \leq a) \quad (70)$$

The first term on the right is seen to be simply the potential of the central ion, k . Consequently, by (60), the second term, which is nothing but the constant A , gives ψ_i , the potential due to the surrounding ion atmosphere. This will be the same everywhere within the region, in particular on the surface of the ion, where $r = b_k$. It is proportional to the charge $Z_k \epsilon$ of the ion. The total potential at the surface of the ion is, of course, obtained by setting $r = b_k$ in equation (70), in which case the first term on the right becomes the self potential of the ion.

With the expression for ψ_i at hand, we may now evaluate the integral (59), which gives the electrical energy of the ion k due to the presence of the surrounding ions. Imagine the charge of the ion to be increased reversibly from zero to its final value, ϵZ_k , by successive increments and let its value at any stage during the charging process be $\lambda \epsilon Z_k$. The process may be regarded as an increase of λ from 0 to 1. As it occurs, the ion atmosphere establishes itself, reversibly, and the resulting potential increases from zero to its final value, ψ_i , given by A in (68). At any intermediate stage, its value is $\lambda \psi_i$. The work is, therefore,

$$W_k = - \int_0^1 \left(\frac{\lambda Z_k \epsilon}{D} \right) \left(\frac{\kappa}{1 + \kappa a} \right) Z_k \epsilon d\lambda = - \frac{Z_k^2 \epsilon^2}{2D} \left(\frac{\kappa}{1 + \kappa a} \right) \quad (71)$$

In this equation Z_k is the only quantity which is peculiar to the k th ion; κ , as will be seen from its definition (65), depends on the composition of the solution as a whole. It is proportional to $\sqrt{\Sigma n_i Z_i^2}$ and has the dimensions of a reciprocal length.

The quantity $\Sigma n_i Z_i^2$ is of key importance in determining W_k , the work of charging an ion. This accords with the observations of G. N. Lewis, which antedated the development of the Debye-Hückel theory. Lewis pointed out that a decisive role in the determination of the activities of electrolytes (as well as other substances) in solution was played by a quantity which he called the ionic strength and which we have already defined in equation (52) as

$$\omega = \frac{1}{2} \Sigma C_i Z_i^2$$

where C_i is the concentration of the i th ion in moles per liter. Except for a numerical factor, this ω is the same as $\Sigma n_i Z_i^2$. By choosing the unit length as the centimeter,

$$\Sigma n_i Z_i^2 = \frac{N}{1000} \Sigma C_i Z_i^2 = \frac{2N}{1000} \omega \quad (72)$$

where N is Avogadro's number.

We may note that the *total* electrostatic free energy of a charged spherical ion of radius b is given from (17), (70), and (71) as

$$W_{\text{total}} = (\bar{F}_e)_{\text{total}} = \frac{Z_k^2 \epsilon^2}{2D} \left(\frac{1}{b} - \frac{\kappa}{1 + \kappa a} \right) \quad (72.1)$$

This formula will be found useful later (Chapters 9 and 11) in dealing with the interaction of ions with polyvalent macromolecules, if the latter can be treated as spheres. For present purposes, however, we are con-

cerned only with the second term in the parentheses of (72.1), since the first is independent of the ionic strength.

We are now in a position to express W , the total work of charging all the ions of a strong electrolyte molecule due to the presence of the surrounding ions. For simplicity we limit ourselves to the case of a binary electrolyte, such as sodium chloride, magnesium sulfate, or sodium sulfate, which gives rise to only two kinds of ions. We suppose that the number of positive ions derived from one molecule of the electrolyte is ν_+ and that their valence is Z_+ , and that the number of negative ions is ν_- and their valence Z_- . Then the expression for W is, from (71),

$$W = \frac{-(\nu_+ Z_+^2 + \nu_- Z_-^2) \epsilon^2}{2D} \left(\frac{\kappa}{1 + \kappa a} \right) \quad (73)$$

and this by (58) is equal to $kT \ln f$. Although f , the activity coefficient of the strong electrolyte molecule, is the quantity we set out to evaluate, it may be conveniently replaced by the mean activity coefficient of the ions, f_{\pm} , which is defined by

$$f_{\pm} = f^{1/\nu}$$

where ν is the total number of ions. When this is done, we arrive at the result

$$\ln f_{\pm} = - \frac{(\nu_+ Z_+^2 + \nu_- Z_-^2) \epsilon^2}{2kTD\nu} \left(\frac{\kappa}{1 + \kappa a} \right) \quad (74)$$

We may simplify this expression by making use of a relation already employed (see the discussion of equation 37):

$$\nu_+ Z_+^2 + \nu_- Z_-^2 = -\nu Z_+ Z_-$$

Then (74) may be written

$$\ln f_{\pm} = Z_+ Z_- \left(\frac{\epsilon^2}{2DkT} \right) \left(\frac{\kappa}{1 + \kappa a} \right) \quad (75)$$

We note that, since Z_+ and Z_- are signed quantities, the product $Z_+ Z_-$ is always a negative integer. This is a common form of the Debye-Hückel equation for strong electrolytes, but it is often written alternatively in terms of ionic strength as

$$\log_{10} f_{\pm} = \frac{Z_+ Z_-}{2.303} \frac{P \sqrt{\omega}}{1 + Qa \sqrt{\omega}} \quad (76)$$

where

$$P = \frac{\epsilon^3}{2.303(DkT)^{3/2}} \sqrt{\frac{2\pi N}{1000}} \quad (76.1)$$

and

$$Q = \sqrt{\frac{4\pi\epsilon^2}{DkT} \cdot \frac{2N}{1000}} \quad (76.2)$$

The transition from (75) to (76) follows at once from the definition of κ and ω . Equation (75) is, of course, limited to a binary electrolyte. (It would be a straightforward matter to adapt it to the more general case of any electrolyte, but the resulting expression is cumbersome. Full discussions are given in numerous references; see for instance Harned and Owen (1950), Robinson and Stokes (1955) and some of the general references given at the end of Chapter 4—for instance, Lewis and Randall, Glasstone, or Rossini.)

It may be of interest to record here values¹¹ of the two constants P and Q for water, the most important solvent for the biochemist, at 25°.

$$P = 0.5045 \quad (\text{water at } 25^\circ)$$

$$Q = 3.286 \times 10^7 \quad (\text{water at } 25^\circ)$$

It will be noted that the value of P in (76.1) is proportional to the product of the dielectric constant and temperature, raised to the $-\frac{3}{2}$ power: $(DT)^{-\frac{3}{2}}$. Likewise the coefficient Q in the denominator in (76) is proportional to $(DT)^{-\frac{1}{2}}$. It is well known that the dielectric constant of any liquid decreases with rising temperature, for reasons that will become apparent in Chapter 6, in the discussion of the theory of the dielectric constant. For water the variation is such that the product DT remains nearly constant from 0° to 60°. Consequently over this entire range it is possible to write equation (76), with an accuracy sufficient for most purposes, at ionic strengths up to 0.3, or thereabouts, by the numerical formula

$$\text{Log}_{10} f_{\pm} = \frac{(Z_+Z_-)0.50 \sqrt{\omega}}{1 + 0.33a' \sqrt{\omega}} \quad (76.3)$$

Here we have set $a' = 10^8 a$, so that a' is expressed in angstrom units. If a' is taken as 3 Å this simplifies still further:

$$\log_{10} f_{\pm} = \frac{(Z_+Z_-)0.50 \sqrt{\omega}}{1 + \sqrt{\omega}} \quad (76.4)$$

and as this is very often a reasonable value to take for a' this simple formula is useful for many calculations. It must be noted that the nu-

¹¹ These are based on the values: $D = 78.54$ (water at 25°); $25^\circ = 298.16^\circ \text{ K}$; $\epsilon = 4.802 \times 10^{-10}$ esu; $k = 1.380 \times 10^{-16}$ erg/deg; and $N = 6.024 \times 10^{23}$. Values of P and Q for other solvents and other temperatures can be calculated from the definitions (76.1) and (76.2).

merical values given in (76.3) and (76.4) are applicable only when the solvent is pure water. If we are dealing with some other solvent with different dielectric constant, the appropriate values of P and Q must be calculated directly from equations (76.1) and (76.2).

At high ionic strengths—of the order of 0.3 to 0.5 M and above—it is commonly found that (76.3) and (76.4) are inadequate to describe the facts. The experimental values of $\log f_{\pm}$ are often found to pass through a minimum, and then increase again at high ionic strengths, so that $\log f_{\pm}$ may even become positive. The curves vary greatly from one salt to another at high ionic strengths; in the alkali halides, for instance, this upward trend in $\log f_{\pm}$ is most pronounced in LiCl , least in CsCl . The simplest extension of (76.3) or (76.4) which approximately fits the facts is the inclusion of a term linear in ω , giving

$$\log_{10} f_{\pm} = \frac{(Z_+Z_-)0.50 \sqrt{\omega}}{1 + 0.33a \sqrt{\omega}} + K_s \omega \quad (76.5)$$

The second term on the right is of opposite sign to the first, and K_s may be looked on as a "salting-out" term. It is chosen empirically for each salt so as to give the best fit to the experimental curve. Although K_s is to be regarded here primarily as an empirical parameter, it is found, for ions containing large nonpolar organic side chains, that K_s is large; and there is clearly a general relation between the K_s term in (76.5) for ions and the salting-out coefficients for related uncharged molecules.

Other, more elaborate equations to describe the activity coefficients of electrolytes at high concentrations have been developed. For instance Robinson and Stokes have given an equation involving hydration factors for the ions, which fits the data for many electrolytes up to quite high concentration. (See the book by Robinson and Stokes, 1955.)

At this point a few general remarks about the Debye-Hückel theory are in order. In the first place, it is worth considering more closely the significance of the quantity κ , which depends only on the total ionic composition of the solution and is to be interpreted as a reciprocal length. If we introduce the value of ψ given by (69) into the approximate equation for the density of charge in the ion atmosphere surrounding the k th ion (63) and employ equation (65), we obtain the result

$$\rho = - \frac{Z_k \epsilon^3}{DkT} \sum n_i Z_i^2 \left(\frac{e^{\kappa a}}{1 + \kappa a} \right) \left(\frac{e^{-\kappa r}}{r} \right) = - \frac{Z_k \epsilon \kappa^2}{4\pi} \left(\frac{e^{\kappa a}}{1 + \kappa a} \right) \frac{e^{-\kappa r}}{r} \quad (77)$$

This expression is valid everywhere outside the sphere of radius $r = a$. It will be seen from it how rapidly ρ drops with an increase in r . But now let us consider not the charge density itself but instead the total charge

contained in a spherical shell of radius r and thickness dr surrounding the ion under consideration. This will be

$$dq = 4\pi r^2 \rho \, dr$$

The quantity $dq/dr = 4\pi r^2 \rho$, which gives the total charge in the shell *per unit thickness*, has the value

$$\frac{dq}{dr} = -Z_k \epsilon \kappa^2 \left(\frac{e^{\kappa a}}{1 + \kappa a} \right) r e^{-\kappa r} \quad (78)$$

This is to be interpreted as the excess of negative over positive charge at any distance r from the k th ion. This quantity, instead of decreasing uniformly like ρ as we depart from the surface $r = a$, passes through a maximum at $r = 1/\kappa$, and the value of this maximum is

$$\left(\frac{dq}{dr} \right)_{\max} = -Z_k \epsilon \kappa \frac{e^{\kappa a - 1}}{1 + \kappa a} \quad (79)$$

As κ goes to zero, the distance at which this maximum occurs recedes to infinity and at the same time the value of the maximum goes to zero. Plots of dq/dr , corresponding to several ionic strengths, are shown in Fig. 12. It will be seen from this what an important parameter κ is in determining the electrical characteristics of the system.

We may note that the total charge $q = \int_a^\infty 4\pi r^2 \rho \, dr$ in the ion atmosphere around the central ion must be equal and opposite to the charge on the ion; that is, it is equal to $-Z_k \epsilon$, since the solution as a whole is electrically neutral. This relation may be readily proved by integration of (78).

The conceptions underlying the definition of the charge density, ρ , should be carefully considered. The value of ρ represents the *average* value of the charge density surrounding a given ion, of charge $Z_k \epsilon$, which is chosen as the center of our coordinate system. If we imagine an observer—a sort of subatomic demon—seated in the center of the ion, watching other ions in the neighborhood come and go, he will notice that ions of opposite charge to his own turn up more frequently than those of the same sign of charge, and that this difference between anions and cations in any volume element is more pronounced, the nearer the volume element is to him. *At any particular instant*, however, the charge density in any neighboring volume element—if the demon makes an instantaneous count of the anions and cations present—will in general be different from the average charge density, $\rho(r)$, since there are constant fluctuations as ions move into or out of the volume element. A very long series of such counts, however, made over successive intervals of time, will give an average

which cancels out the influence of the positive and negative fluctuations. The fact that ρ represents an average value is a natural consequence of the fact that we have used the Boltzmann distribution law to calculate the relative numbers of ions having different electrical energies; for the Boltzmann law is a statistical law, derived by statistical reasoning.

The coordinate system of the demon inside the ion is of course in constant random motion, if we adopt the point of view of another similar

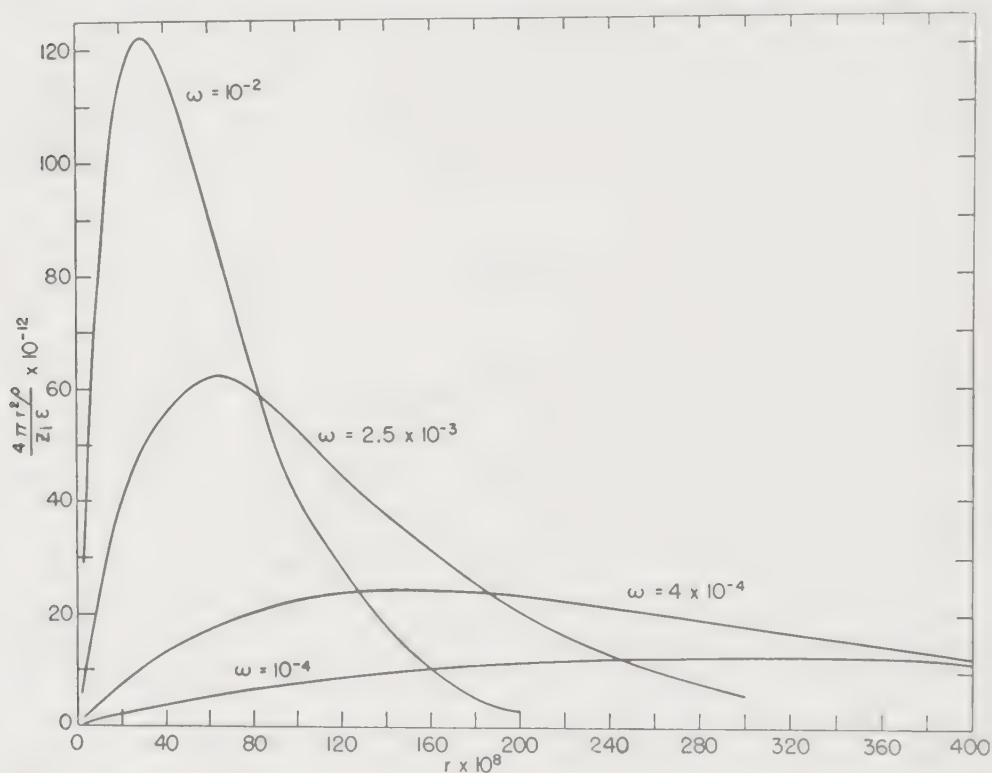


FIG. 12. The distribution of charge around the central ion as a function of distance, taking the collision diameter as 3×10^{-8} cm (3 Å). Note that the area under all the curves in this figure must be the same, if the curves are extended out to infinite distance, since the total surrounding charge must be equal and opposite to the charge on the central ion. The maximum in the distribution curve occurs at the distance $1/\kappa$ from the central ion.

observing demon who is ensconced in a niche on the walls of the container, looking out into the solution. If the walls are uncharged, the second demon, looking out onto any neighboring volume element of solution, will also notice rapid fluctuations in the charge density in that volume element from instant to instant. If he takes repeated counts and averages them over a sufficient period of time, however, he will find that the average charge density, in any volume element he may have selected to observe, always turns out to be zero. This is inevitable, since there is

nothing in this frame of reference which favors either anions or cations.¹² On the other hand, the demon inside the ion taken as center is operating from a system which, because of the charge on the ion, has a built-in bias in favor of ions of opposite charge. It is this coordinate system, centered in the ion and moving with it in its random travels through the solution, which is consistently employed in the development of the Debye-Hückel theory.

Another point concerns the charging process we have invoked as a fundamental part of the calculation of the electrical energy of an ion. This is known as the Güntelberg-Müller process. According to it only the ion under consideration is supposed to receive a charge, by infinitesimal increments. The surrounding ions are assumed to be already charged, and the only change which they suffer during the process is a gradual redistribution as the central ion receives its charge. In the original theory of Debye, an alternative, and perhaps more realistic, charging process is invoked in which all the ions are charged simultaneously. The process leads to an expression for the total electrical energy of the system in terms of the number of molecules or moles of each component present. By differentiating it partially with respect to the number of moles of a given component it is possible to obtain the chemical potential, and indirectly the activity coefficient, of that component in so far as it depends on electrostatic factors. This rather more elaborate and sophisticated approach makes it possible to take account of volume changes and changes of dielectric constant associated with the presence of a given electrolyte. (A good discussion is given by Scatchard in Chapter 3 of Cohn and Edsall, 1943.) In our treatment we have assumed the dielectric constant of the ionic solution to be the same as that of the pure solvent. For moderately dilute solutions, this is probably not too far from the truth and in any case, in view of other approximations involved in the theory, it need not occasion too much concern.

A final word is required regarding the approximation of replacing the exponentials of the Boltzmann distribution terms by their series expansions. This limits the applicability of the theory to cases where the ratio of electrical to thermal energy ($Ze\psi/kT$) is small. In the case of solvents of low dielectric constant and of small and highly charged ions where the

¹² We assume in this discussion that the wall is uncharged. Many actual surfaces, of course, are charged. In the neighborhood of a charged wall the ions will distribute themselves unevenly, ρ being of opposite sign to the charge on the wall, and falling away rapidly with increasing distance from the wall. Calculations of charge distribution in a system of this sort can readily be made by the reasoning of the Debye-Hückel theory and form the basis for modern theories of the electric double layer at charged surfaces.

electrical energy will be high, we must be prepared, therefore, for deviations from the theory. Various attempts to overcome the shortcomings due to the approximation of replacing the exponentials by the first two terms of their series expansions have been made, but we shall not go into these here. It is more instructive to consider the application of the theory to various problems of chemical and biochemical interest, some of which are discussed later in this chapter and others in later chapters, especially those dealing with acid-base equilibrium (Chapters 8 and 9) and with molecular interactions (Chapter 11).

Ion-Dipole Interactions

In the previous section, we have dealt with the activity coefficient of ions present in an ionic environment of sufficient concentration so that electrostatic interactions are important. Of equal, or even greater, interest to the biochemist is the question of the activity of dipolar ions under similar conditions. It is to this that we shall devote ourselves in the present section. It is to be expected at the start, owing to the great difference between the field surrounding a dipolar ion and that surrounding a simple ion, that the results will be quite different and, on the whole, more complex. We need not be surprised, therefore, to find, as we shall in the following pages, that when we consider the effect of surrounding ions (neglecting dipole-dipole interactions) the logarithm of the activity coefficient of a dipolar ion at low ionic strength is proportional to the ionic strength itself, rather than to its square root as in the case of simple ions.

It is helpful to begin with a general picture which indicates the situation in a qualitative way. The simplest dipolar ion is glycine ($+H\text{N}_3^+\text{CH}_2\cdot\text{COO}^-$); if we make a space model, using the bond distances and bond angles given in Chapter 3, we find that the distance from the positively charged nitrogen atom to a point midway between the two oxygens of the negatively charged carboxyl group is slightly over $3 \text{ \AA} = 3 \times 10^{-8} \text{ cm}$. This may be called the dipole distance, R , assuming that our simplified picture of the molecule suffices to define the centers of positive and negative charge as being located in these two positions. The dipole moment, μ , is equal to R multiplied by the proton charge, e ; hence $\mu \cong 3 \times 10^{-8} \times 4.8 \times 10^{-10} \cong 14.4 \times 10^{-18} \text{ esu} = 14.4$ Debye units. The space model shows that glycine may be roughly represented by a sphere of radius approximately 2.8 \AA , with the two charges placed on a diameter, and equidistant from the center of the sphere. An α -amino acid such as leucine, with a large hydrocarbon side chain, should have a dipole distance nearly the same as that of glycine, but in this case the dipole is off to one side of the molecule, which might be roughly

represented by an ellipsoid with the dipole at one focus. An amino acid such as ϵ -aminocaproic acid, $^+\text{H}_3\text{N}(\text{CH}_2)_5\text{COO}^-$, may be expected to be (and indeed is) characterized by a considerably larger dipole distance and moment than an α -amino acid; however, rotation around the valence bonds in the main chain permits a variety of configurations of differing dipole moments, from a fully extended chain with maximum moment, at one extreme, to a more or less ringlike structure, with the $-\text{NH}_3^+$ and $-\text{COO}^-$ groups very close together, at the other. The most probable configuration seems to be something intermediate, as we shall see particularly from the dielectric constant studies recorded in Chapter 6; a suitable simple model might be an ellipsoid of revolution, with a charge of $+\epsilon$ at one focus and of $-\epsilon$ at the other. Long-chain polypeptides composed of amino acids with uncharged side chains are qualitatively similar, with one N-terminal ammonium group and one C-terminal $-\text{COO}^-$ group. Such peptide chains may be more or less randomly coiled, or they may be regularly coiled into α -helices or other helical configurations, as we have discussed in Chapter 3. In the helical structures the dipole moments should be very large, not only because of the positive and negative charges at opposite ends of a rod-shaped molecule, but because the highly polar $\text{C}=\text{O}$ and $\text{N}-\text{H}$ groups of the peptide links are all lined up nearly parallel to the helical axis (see Chapter 3, Fig. 12). Finally, peptides with charged side chains, and proteins, offer innumerable possibilities for the arrangement of many positively and negatively charged groups in complicated structures of almost any size or shape.

To form a simple picture of a dipolar ion surrounded by an ion atmosphere, consider first an ellipsoid of revolution, with a charge $+\epsilon$ at one focus, and $-\epsilon$ at the other, separated by the distance R (Fig. 13). We take the origin of our system of coordinates at the center of the ellipsoid; the polar axis (Z axis) is parallel to the dipole moment, the positive end of the dipole lying on the positive Z axis. Let r be a vector drawn from the origin to the center of an ion, at distance r from the center of the dipolar ion. Then the potential at r due to the dipolar ion is given by equation (23), if r is large compared to the dipole distance, R :

$$\psi_{\text{dipolar ion}} = \frac{\mu \cos \theta}{Dr^2} = \frac{\epsilon R \cos \theta}{Dr^2}$$

Here D , as usual, is the dielectric constant of the medium, and θ is defined by the drawing in Fig. 13. It is clear, by the same type of argument used in developing the Debye-Hückel theory, that the space charge due to the ion atmosphere will be negative on the right-hand side of the diagram (nearer to the positive end of the dipole) and positive on the left-hand side (nearer to the negative end of the dipole). It becomes zero in

the plane $Z = 0$, passing through the origin and perpendicular to the plane of the paper in Fig. 13, since every point in this plane is equidistant from the two ends of the dipole. Qualitatively the effect of the resulting atmosphere is the same for a dipolar ion as for an ion; the potential due to the ion atmosphere is everywhere opposite in sign to the potential due to the dipole itself. Therefore the ion atmosphere lowers the electrical free energy of the dipolar ion, and hence lowers its activity coefficient also. However, the charge density in the ion atmosphere around a dipolar ion falls off much more rapidly with increasing distance than that around an ion, since the potential around the dipolar ion varies (for constant θ) as $1/r^2$ instead of $1/r$. This is the fundamental reason

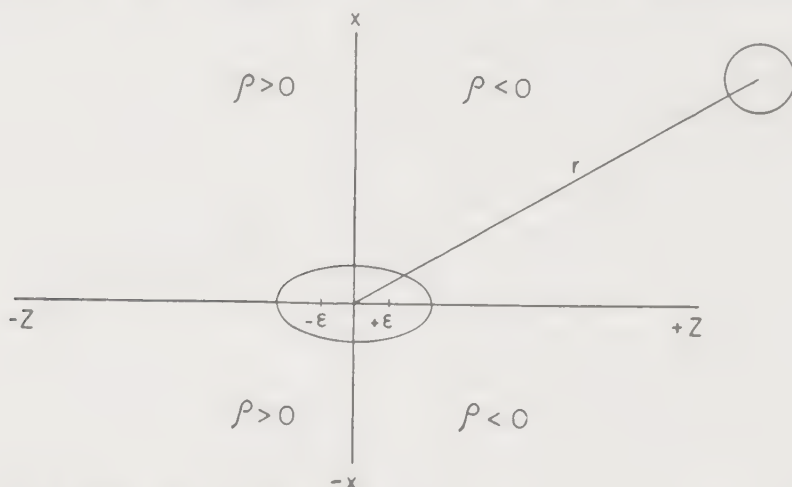


FIG. 13. A simple model of an ellipsoidal dipolar ion and a neighboring ion (see text). The angle θ is defined as the angle between the positive z axis and the vector \mathbf{r} .

why $-\log f$, for a dipolar ion, when plotted as a function of ionic strength (ω) contains no term proportional to $\sqrt{\omega}$; the limiting law, at low ionic strength, is always of the form $-\log \gamma = K_R \omega$, where K_R is a constant. We now consider how to convert this general picture into quantitative terms.

In dealing with the general problem of reckoning the electrostatic energy of a dipolar ion in the presence of an ionic and dipolar ionic environment the natural thing to do would be to follow the lines of the Debye-Hückel theory, reckoning the potential surrounding the dipolar ion and then carrying out a Güntelberg-Müller charging process on it. This would involve expressing the distribution of both simple ions and dipolar ions round the dipolar ion in question by means of the Boltzmann distribution law in terms of the electrostatic energy of each, and then making use of the result in connection with Poisson's and Laplace's equations. Such a procedure presents severe mathematical difficulties

which will be appreciated from the relatively complicated form of the expression for the potential of even a point dipole. To be sure, the process could be much simplified by neglecting the effect of surrounding dipoles, but even so it would be formidable. Kirkwood has modified the approach and given a somewhat different treatment of the whole problem, which he applies to solutions sufficiently dilute in dipolar ions so that dipole-dipole interactions may be neglected in comparison with ion-dipole interactions. This treatment has the advantage that it may be used in connection with models which make it possible to take account of effects not contemplated in the original Debye-Hückel treatment, namely salting-out effects.

The first stage of the Kirkwood treatment consists in the development of an alternative to the more complete and rigorous statistical procedure of the Debye-Hückel theory. It leads directly to an approximate expression for the work of charging a dipolar ion in the presence of a surrounding ion atmosphere, that is, for the work of a Güntelberg-Müller charging process, valid in the limiting case of an infinitely dilute solution. This expression is a general one and may in principle be applied to any model of a dipolar ion, subject only to the conditions that the integrals involved converge. (Indeed, were it not for this condition, it would be applicable to the charging of a simple ion, but in this case it turns out that the integrals become infinite.) The second stage of the treatment consists in the application of this expression to three special models, none of which would be amenable to more rigorous treatment along the lines of the original Debye-Hückel theory. Each of these models represents a compromise between what can be handled mathematically and what gives a reasonable picture of an actual type of dipolar ion. They are, respectively: (1) a sphere with a point dipole at its center; (2) a prolate ellipsoid of revolution with a positive charge at one focus and a negative charge at the other; (3) a prolate ellipsoid of revolution with a point dipole at one focus. In each of these the dipolar ion—sphere or ellipsoid—is treated as a cavity of low dielectric constant embedded in a continuum of high dielectric constant in which the surrounding ions are distributed. Although in Kirkwood's full expression for the work of charging a term is included to take account of dipole-dipole interactions, this is not made use of in practice, for reasons of mathematical simplicity, and it is assumed that the concentration of dipolar ions is negligible in comparison with that of ordinary ions.

The expression for the work of the charging process (neglecting the effect of surrounding dipolar ions) may be derived as follows. Consider any dipolar ion in the presence of m species of simple ions. Let the average concentration of ions of species i , for the solution as a whole, be n_i .

Introduce any convenient system of coordinates fixed with reference to the dipolar ion (Fig. 13). Let the energy of an ion of species i at a given point, and, therefore, in a given fixed configuration with respect to the dipolar ion, be W_i .¹³ The energy W_i will result, of course, from the total constellation of ions surrounding the dipolar ion as well as from the dipolar ion itself. If the charges and, therefore, the moment of the dipolar ion were some fraction, λ , of their full value, the state of affairs would be different and W_i would have some other value, $W_i(\lambda)$. The Boltzmann distribution law tells us that for any value of λ the number of molecules of species i contained in an element of volume dv is

$$dn_i = n_i e^{-W_i(\lambda)/kT} dv \quad (80)$$

Now suppose that the work of bringing an ion, i , from an infinite distance into the element of volume, due solely to the presence of the dipolar ion (i.e., in the absence of all other ions), is V_i when the dipolar ion has its full charge, i.e., its full moment. Then, when the moment is a fraction of its full value, the corresponding work is λV_i . The work of increasing the moment of the dipolar ion by an amount $d\lambda$ in the presence of all the ions in the element of volume dv is therefore

$$\sum_{i=1}^m \int_{\lambda=0}^{\lambda=1} n_i V_i e^{-W_i(\lambda)/kT} dv d\lambda$$

In order to obtain the total effect of *all* the surrounding ions, we must integrate this over the whole volume of the solution available to the ions. This is the total volume except for the region from which they are excluded by the presence of the dipolar ion itself. We shall denote this region by χ . The total work of establishing the charges (moment) of the dipolar ion is therefore

$$W = \sum_{i=1}^m n_i \int_{v=\chi}^{\infty} \int_{\lambda=0}^1 V_i e^{-W_i(\lambda)/kT} dv d\lambda \quad (81)$$

For simplicity we have taken the volume of solution as infinite, instead of having some finite value, since the effect of the surrounding ion atmosphere falls off rapidly with distance. The W of equation (81) is what determines the activity coefficient, f , of the dipolar ion, which is given by

$$kT \ln f = W \quad (82)$$

¹³ We do not have to consider the orientation of an ion, (as we do in the case of a dipolar ion) but only its position.

In order to implement equation (81) it is necessary to know V_i as a function of the space coordinates and $W_i(\lambda)$ as a function both of the space coordinates and of λ . The function $W_i(\lambda)$ presents quite a problem, for, as we have pointed out, it depends not only on the dipolar ion under consideration but on the full constellation of surrounding ions to which it gives rise. To meet this difficulty, we introduce an approximation which is strictly valid only in the limiting case of solutions in which the concentration of ions, though large in comparison with that of the dipolar ions, itself approaches zero. This consists in setting

$$W_i(\lambda) = \lambda V_i \quad (83)$$

It means that we neglect the effect of surrounding ions in the work of bringing a given ion up to any specified distance from the dipolar ion under consideration. When this value of $W_i(\lambda)$ is introduced into (81), the integration with respect to λ may be carried out quite simply, with the result

$$W = \sum_{i=1}^m n_i k T \int_{v=x}^{\infty} (1 - e^{-V_i/kT}) dv \quad (84)$$

If we are content with the first three terms of the series expansion of the exponential, this becomes

$$W = \sum_{i=1}^m n_i k T \int_{v=x}^{\infty} \left[\frac{V_i}{kT} - \frac{1}{2} \left(\frac{V_i}{kT} \right)^2 \right] dv \quad (85)$$

Save for the omission of a term representing dipole-dipole interaction, equation (84) is the general expression for the work of a Güntelberg-Müller charging process applied to a dipolar ion in very dilute solution. The great merit of this expression, and of the whole treatment, is that it reduces the problem to that of calculating V_i , the mutual energy of only two molecules, a dipole and a surrounding ion, as a function of the space coordinates. This is an enormous simplification.

It now remains, as the second and final stage of the analysis, to apply this result, embodied in equation (84), to each of the three special models described above. The principal task in this connection is that of working out the expressions for V_i , the integration itself being relatively simple. This is a somewhat arduous process, even for the simplest of the models, and involves finding solutions of Laplace's equation fitted to the singularities resulting from the special distribution of charges assumed and the boundary conditions. These are the continuity of the potential and of

the normal component of the electric displacement¹⁴ at the surface of the cavity representing the dipolar ion, where the dielectric constant is assumed to jump from its low value characteristic of the interior of the ion to its macroscopic value. We shall not attempt to give Kirkwood's calculations in detail, but shall content ourselves with summarizing his results for each of the three models. Before doing so, however, in order to convey some idea of the kind of procedure involved and of the way in which the method works, we shall analyze in detail a much simpler model. The results which emerge are not without interest in their own right.

The principal complications in the application of (84) to the Kirkwood models result from treating the dipolar ion as a cavity of low dielectric constant. As a simpler model, therefore, we shall regard the dipolar ion as a point dipole of moment μ embedded in a continuous medium of

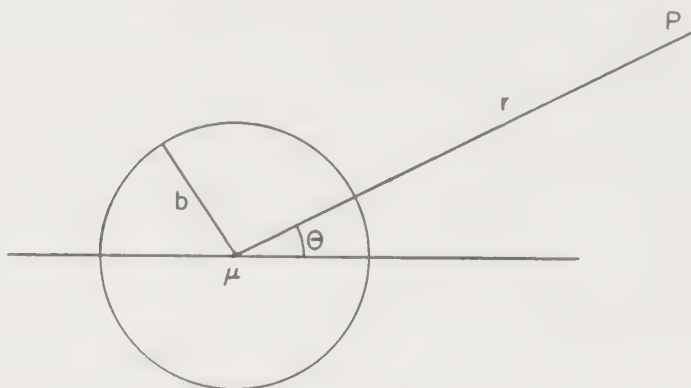


FIG. 14. The system of coordinates used by Kirkwood in calculating the interaction between ions and dipolar ions.

dielectric constant D . We employ a system of polar coordinates with origin in the dipole and polar axis parallel to the dipole axis. We consider only the interaction between the dipole and the surrounding i species of ions. We suppose that the closest approach of an ion to the dipole is given by a spherical surface of radius b (Fig. 14). Then the potential of the dipole, which, for reasons of symmetry, is independent of ϕ , is given at any point (r, θ) by equation 23:

$$\psi = \frac{\mu \cos \theta}{Dr^2}$$

The work of bringing up an ion bearing the charge $Z_i e$ to this point is therefore

$$\frac{Z_i e \mu \cos \theta}{Dr^2} \quad (86)$$

¹⁴ The electric displacement is equal to DE , the product of dielectric constant and electric intensity. It is further discussed in Chapter 6.

This is the value of V_i in equation (85). The element of volume dV may be expressed in terms of θ and r as $2\pi \sin \theta r^2 d\theta dr$ and the integration is over the range from $\theta = 0$ to $\theta = \pi$, and from $r = b$ to $r = \infty$. Since

$$\int_0^\pi \cos \theta \sin \theta d\theta = 0$$

the first term of the integral $\left[\int_x^\infty (V_i/kT) dv \right]$ vanishes, and we are left with

$$\begin{aligned} W &= kT \log f \\ &= \sum_{i=1}^m n_i kT \int_{\theta=0}^\pi \int_{r=b}^\infty - \left[\frac{1}{2(kT)^2} \left(\frac{Z_i \epsilon \mu \cos \theta}{Dr^2} \right)^2 \right] 2\pi \sin \theta r^2 d\theta dr \end{aligned}$$

The integration may be carried out in a perfectly straightforward manner with the result that

$$\ln f = - \left(\frac{\mu \epsilon}{DkT} \right)^2 \frac{2\pi}{3b} \sum_i n_i Z_i^2$$

or, after introduction of the ionic strength defined by

$$\begin{aligned} \omega &= \frac{1000}{2N} \sum n_i Z_i^2 \\ \log f &= - \left(\frac{\mu \epsilon}{DkT} \right)^2 \left(\frac{2\pi N}{2303} \right) \frac{2\omega}{3b} \end{aligned} \quad (87)$$

This yields the important relation, preserved in the more complex models, that at high dilution the surrounding ion atmosphere acts to decrease the activity coefficient of the dipolar ion and that the effect on $\log f$ is proportional to the first power of the ionic strength, to the square of the dipole moment,¹⁵ and to the inverse square of the dielectric constant and temperature. All these predictions are realized in experiments on the interactions of ions with amino acids and peptides in solvents of different dielectric constants. This very simple treatment, ignoring the microscopic dielectric properties of the dipolar ion, misses, of course, the salting-out effect which enters only when we treat the molecule as a cavity of low dielectric constant in the medium of relatively high dielectric constant.

We may now summarize the results of Kirkwood's analysis of each of his three models.

¹⁵ In one of the models discussed below (prolate ellipsoid with charges at the foci) $\log f$ is proportional to the moment itself, rather than its square.

Model 1. Here the dipolar ion is treated as a sphere of radius b with a point dipole of moment μ at its center. The distance of closest approach of the dipolar ion and a surrounding ion is denoted by a . The limiting law valid for infinite dilution may be written as

$$\log_{10} f = -K\omega$$

where ω denotes ionic strength and

$$K = \frac{2\pi N\epsilon^2}{2303DkT} \left[\frac{3}{2} \frac{\mu^2}{DakT} - \frac{b^3\alpha(\rho)}{a} \right] = - \left(\frac{\partial \log f}{\partial \omega} \right)_{\omega=0} \quad (88)$$

Here ρ is the ratio b/a and $\alpha(\rho)$ is a function of ρ , the value of which is 1 when $\rho = 0$, and 1.96 when $\rho = 0.9$. The most striking feature of this result is that, in contrast to the case of simple ions, the activity coefficient of a dipolar ion depends on the first power—not the one-half power of the ionic strength. We have already referred to this. Another important point is that the over-all constant K is decomposable into two parts. The first, proportional to $(\mu/DT)^2$, results from the electrostatic interaction of the dipole and the surrounding ions and corresponds to the decrease in activity coefficient due to the ion atmosphere. The second, which is proportional to $(DT)^{-1}$, does not depend on the moment of the dipolar ion, but only on its geometrical properties as a cavity of low dielectric constant in which image charges are produced by the surrounding ions. It corresponds to a salting-out effect, which may be thought of in terms of a repulsion of the ions and their image charges. It is to be noted that the dielectric constant of the interior of the cavity is lost in the calculations. This results from its being neglected in relation to the much larger macroscopic dielectric constant. Notwithstanding this, the procedure of treating the dipolar ion as a cavity of low dielectric constant is a major feature of Kirkwood's treatment, since it is this which gives rise to the salting-out term. No such effect is provided for in the model underlying the classical Debye-Hückel theory when the internal dielectric constant of an ion is not taken into consideration.

Model 2. Here the dipolar ion is treated as a prolate ellipsoid of revolution with a charge $+\epsilon$ at one focus and $-\epsilon$ at the other. For this case again

$$\log_{10} f = -K\omega$$

Here K is given by

$$K = \frac{2\pi N\epsilon^4 g(\lambda_0) R}{2303(DkT)^2} = - \left(\frac{\partial \log f}{\partial \omega} \right)_{\omega=0} \quad (89)$$

where R is the distance between the foci, and g is a function of λ_0 , the reciprocal of the eccentricity of the ellipsoid. For most ellipsoids g lies

between 0.5 and unity. It will be seen that here K is proportional to the first power of R , and hence of the moment, but again inversely to the square of the dielectric constant. No salting-out term was included in the calculations for this model.

Model 3. Here the dipolar ion is treated as a prolate ellipsoid of revolution with a point dipole of moment μ , parallel to the major axis, present at one focus. For this case also

$$\log_{10} f = -K\omega$$

In this case

$$K = \frac{3\pi N\epsilon^2 h(\lambda_0)}{2303(DkT)^2 R} \frac{\mu^2}{R} = \left(\frac{\partial \log f}{\partial \omega} \right)_{\omega=0} \quad (90)$$

where once more R is the distance between foci, and $h(\lambda_0)$ is a function of the eccentricity of the ellipsoid. In this case, as in the last, the salting-out term has not been retained.

All these models of dipolar ions contain only a single positive and a single negative charge. Kirkwood (1934) has given a much more general treatment, however, for more complicated distributions of many charges within a molecule, including ions carrying a net charge as well as dipolar ions. The treatment is restricted to spherical models, but it applies to any arbitrary distribution of charged groups within or upon the surface of the sphere. The mathematical theory is straightforward, but more elaborate than that which we have discussed here. Since multivalent peptides and proteins contain numerous charged groups located in different parts of the molecule, it is plain that any adequate theory of the properties of macromolecular ions and dipolar ions must deal with such complicated charge distributions. Up to the present no very systematic attempt to deal with the theory of such complicated systems has been made. Some discussion of the solubility of proteins in dilute salt solutions, using simplified models derived from the Kirkwood theory for spherical ions, has been given by Cohn and Ferry (see Cohn and Edsall, 1943, pp. 609-616). Now that advances are being made in the detailed knowledge of the structure of proteins (see Chapter 3), it may be expected that the theory for model substances containing many charges will assume an added interest. The generalization of the treatment to include ellipsoidal as well as spherical models, containing an arbitrary charge distribution, would greatly increase the power of the treatment. Some important calculations for spherical and ellipsoidal models have been given in a recent paper by Linderstrøm-Lang (1953).

We now turn from these theoretical considerations to a study of some of the experimental data on the properties of dipolar ions.

Experimental Studies on the Solubility of Dipolar Ions, and Their Interaction with Ions

Numerous experimental studies on the properties of dipolar ions have been carried out, and the results are discussed in detail by Cohn and Edsall (1943). For studies of the interaction of ions and dipolar ions, solubility measurements are of particular value, and we shall confine our discussion here to them, in comparing the experimental data with Kirkwood's theoretical equation. Several other types of thermodynamic measurements, however, such as electromotive force measurements and freezing point depressions, have also been employed in characterizing these

TABLE II
SOLUBILITIES OF SOME DIPOLAR IONS AND UNCHARGED ISOMERS IN WATER AND ETHANOL AT 25°

Substance	Solubility (moles/liter)		Solubility (log <i>N</i>)		$\Delta \log N$
	In water	In ethanol	In water	In ethanol	
Glycine	2.886	0.00038	-1.247	-4.638	3.391
DL- α -Alanine	1.656	0.00076	-1.491	-4.347	2.856
DL- α -Amino- <i>n</i> -butyric acid	1.800	0.00260	-1.440	-3.818	2.378
DL- α -Amino- <i>n</i> -caproic acid	0.0866	0.00104	-2.801	-4.215	1.414
L-Leucine	0.171	0.00128	-2.503	-4.125	1.622
L-Asparagine	0.186	0.000023	-2.468	-5.870	3.402
β -Alanine	6.123	0.00189	-0.816	-3.955	3.139
ϵ -Aminocaproic acid	3.848	0.00194	-0.975	-3.947	2.972
Diglycine	1.512	2.22×10^{-5}	-1.522	-5.889	4.367
Triglycine	1.0229	1.06×10^{-6}	-2.241	-7.206	4.965
Glycolamide	5.509	0.342	-0.900	-1.699	0.799
Lactamide	8.779	2.847	-0.506	-0.759	0.253

These data are taken from Cohn and Edsall (1943, Chapter 9), where many other data are also given. The value of $\Delta \log N$ denotes the solubility increment in passing from ethanol to water.

interactions. Here we shall attempt to give only an introductory discussion and survey of the experimental results, beginning first with some general comments on the solubility of dipolar ions and their uncharged isomers in various media (see also Edsall, 1947).

Dipolar ions, like ionic salts, are generally far more soluble in water than in media of lower dielectric constant. Some examples are given in Table II. The solubilities are given both in moles of solute per liter of solution, which is a convenient unit for practical purposes, and as the

logarithm of the mole fraction of solute ($\log N$), which is a more suitable unit for most thermodynamic calculations.

Several points are apparent from the data of Table II, and from other data which are available. (1) Comparison of a dipolar ion with an uncharged isomer—for instance glycine ($^+\text{H}_3\text{NCH}_2\text{COO}^-$) and glycolamide ($\text{HOCH}_2\text{CONH}_2$)—shows that the dipolar ion is far more soluble in water, relative to ethanol, than the uncharged isomer. The increment in solubility, $\Delta \log N$, on proceeding from ethanol to water, is 3.40 for glycine, 0.80 for glycolamide. The difference, 2.6, is almost exactly the same as the difference of the corresponding increment in $\log N$ between their respective higher homologs, α -alanine ($\Delta \log N = 2.86$) and lactamide ($\Delta \log N = 0.25$). This suggests—and there is other evidence to confirm the suggestion—that the transformation of an uncharged isomer into a dipolar ion produces a fairly characteristic increment in the water/ethanol solubility ratio. This statement should be taken as referring specifically to dipolar ions with a given separation between the charged groups, in this case with only one carbon atom separating the positively charged amino group from the negatively charged carboxyl group. (2) We may remark that the solubility of glycine in other liquids of still lower dielectric constant is even lower than in ethanol; expressing the data as mole fraction glycine at saturation, the value in *n*-butanol is about one-third that in ethanol, and in acetone it is about one-third that in *n*-butanol. In ether, and still more so in truly nonpolar solvents like benzene or hexane, the solubility of dipolar ions is almost immeasurably low. (3) In a homologous series such as glycine, α -alanine, α -amino-*n*-butyric acid, α -amino-*n*-valeric acid, α -amino-*n*-caproic acid, the value of $\Delta \log N$ decreases approximately 0.50 for each additional CH_2 group inserted in the molecule. This rule is perhaps more easily remembered if stated in terms of the solubility ratios themselves, rather than their logarithms: the ratio (solubility in water)/(solubility in ethanol) decreases approximately threefold per added CH_2 group in a straight-chain compound. If lactamide is compared with glycolamide, it may be seen from the data of Table II that the same rule holds for these uncharged isomers of glycine and alanine. A branched-chain isomer has a slightly higher value for the above ratio than the corresponding straight-chain compound; compare the isomers α -amino-*n*-caproic acid and leucine in Table II. Also when two isomeric amino acids which differ with respect to the number of atoms separating the charged groups are compared—for instance α - and β -alanine, or α - and ϵ -aminocaproic acid—the isomer with the greater charge separation has the higher solubility ratio. (4) The absolute values of solubility in any one solvent show no simple relation to the structures of the individual molecules concerned; increase in length

of a hydrocarbon side chain is generally accompanied by decreasing solubility in water, for instance, but α -amino-*n*-butyric acid is more soluble in water than α -alanine, in contradiction to the general trend. This is not surprising, for the absolute value of solubility depends on the work required to separate the molecules from one another in the crystal lattice as well as on their affinity for the solvent molecules. A high value of the crystal lattice energy in itself tends to make the solubility of the substance low in any solvent. Dipolar ions tend, like salts, to pack into crystals so as to give a high lattice energy, because of the electrostatic attractions between the charged groups. The shapes of the molecules, however, also affect the packing; side chains projecting at an awkward angle, for instance, may hinder the effective juxtaposition of the oppositely charged groups to form a regular pattern in the crystal, so that the energy of binding between the components of the crystal lattice is diminished. This kind of obstruction to close packing may account for the high solubility of α -amino-*n*-butyric acid in water, although we do not yet have direct evidence for the crystal structure of this compound, as we do for glycine and α -alanine (see, for instance, Corey, 1948).

The *relative* solubility of a given substance in two different media, however, as we have seen in the preceding discussion, frequently shows a simple and regular relation to structure.

SOLUBILITIES AND ACTIVITY COEFFICIENTS IN WATER AND OTHER MEDIA

The determination of relative solubility in different solvents provides data for determining the free energy of transfer of the solute from one medium to another. To visualize the relations involved, imagine first an ideal experiment in which the two solvents under consideration are separated by a semipermeable membrane, impermeable to both solvents but permeable to the solute, which is present at infinite dilution. At equilibrium the chemical potential of the solute must be the same in both phases. If the standard state is taken as being the same for both phases, the activity, a , in both must also be the same. We assume that one of the solvent phases is pure water, which is taken as the standard medium. Denoting the mole fraction of solute in the aqueous phase by N_w° and in the other solvent by N_A° , and denoting the corresponding activity coefficients by f_w and f_A , respectively, we have

$$a = f_w N_w^\circ = f_A N_A^\circ \quad (91)$$

In an infinitely dilute aqueous solution, however, $f_w = 1$ for the solute, by the convention which assigns this state as the standard state. Hence

$$f_A = N_w^\circ / N_A^\circ \quad (92)$$

Consider now a saturated solution of the same solute (at mole fraction N_w) in water, and in the other solvent (at mole fraction N_A). *If the same solid phase is in equilibrium with both saturated solutions*, then the activity of the solute, a_s , must be the same in both solutions, since both are in equilibrium with the solid phase:

$$a_s = N_w(f_w)_{\text{sat}} = N_A(f_A)_{\text{sat}} \quad (93)$$

Since the solubility of amino acids, and dipolar ions in general, is very low in most organic solvents, we may generally assume the activity of the solute to be proportional to its concentration in such solvents, at all concentrations up to and including the saturated solution. This means that we may set f_A in (92) equal to $(f_A)_{\text{sat}}$ in (93). Thus the activity coefficient of the solute in any solvent is given by its solubility ratio in water and that solvent, multiplied by its activity coefficient in the saturated aqueous solution.

$$f_A = [N_w/N_A]_{\text{sat}}(f_w)_{\text{sat}} \quad (94)$$

The values of $(f_w)_{\text{sat}}$ in water are generally not very far from unity. Smith and Smith (1937, 1940) determined values of f_w as a function of molality (moles of amino acid per kilogram of water) for a number of amino acids and peptides over a wide range of concentration, in aqueous solution, often up to complete saturation. They employed isopiestic vapor pressure measurements, a method to which we have briefly alluded in Chapter 4. Their results, for a number of saturated (or nearly saturated) solutions have been tabulated by Cohn and Edsall (1943, p. 198).^{*} For glycine in saturated aqueous solution (3.33 moles per kilogram of water or $N = 0.0566$), $\log (f_w)_{\text{sat}}$ is -0.112 . From the data of Table II, for glycine in ethanol and water, $\log [N_w/N_A]_{\text{sat}}$ is 3.391, and therefore, from equation (94) in logarithmic form, $\log f_A = 3.391 - 0.112 = 3.279$. The correction term here, due to $\log (f_w)_{\text{sat}}$, is only a second-order correction, which for other dipolar ions of lower solubility in water may often be neglected in approximate calculations.

The molal free energy of transfer of the solute from aqueous solution (at mole fraction N_w) to the medium A (at mole fraction N_A) is given by the difference in chemical potential of the solute in water (μ_w) and in the medium A (μ_A).

$$\begin{aligned} \mu_A - \mu_w &= RT \ln (a_A/a_w) \\ &= RT \ln (N_A/N_w) + RT \ln f_A - RT \ln f_w \end{aligned} \quad (95)$$

Here of course the solute activity in water $a_w = f_w N_w$, with a corresponding expression for the activity of the solute in the medium A . If

^{*} See also Table II of Chapter 4 in the present volume.

the solutions are both saturated, then $\mu_A = \mu_w$, since both are in equilibrium with the same solid phase, and the chemical potential of the solute is thereby fixed. In this case the whole expression on the right of (95) becomes equal to zero, and (95) becomes equivalent to (94), on taking the logarithms of the terms in (94) and multiplying by RT .

SOLUBILITIES IN THE PRESENCE OF SALT

We now consider the effects of increasing the ionic strength on the solubility of glycine in various ethanol-water mixtures. The general character of the data is shown in Fig. 15; the range of solvents covered is from water to 95% ethanol. The solubilities are expressed as $\log N$, and

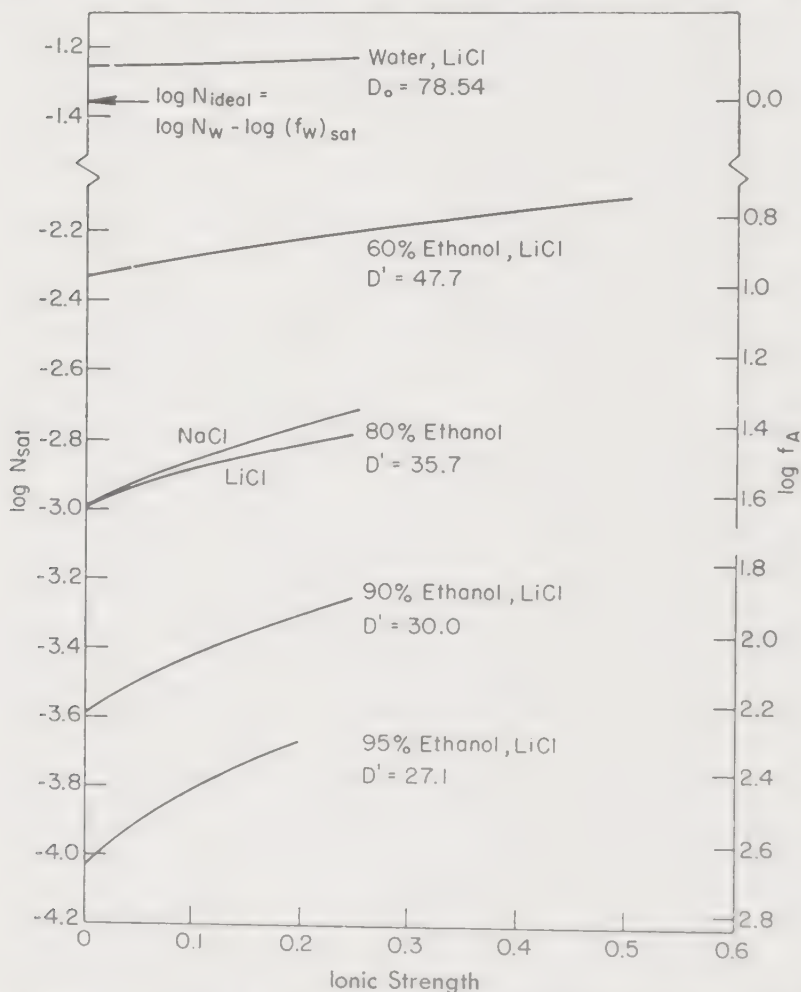


FIG. 15. The solubility of glycine in water, ethanol-water mixtures and salt solutions at 25°. The dielectric constant D' of the pure solvent in the absence of glycine or salt is indicated in each case. Note the progressive decrease of the solubility of glycine at zero ionic strength, and also the increase in the salting-in action of added salt, as D' decreases.

the dielectric constant (D') of each of the pure solvents is indicated. The dielectric constant of water, the standard medium, is denoted by D_0 . There are obviously two major effects of decreasing the dielectric constant of the solvent: (1) In the absence of added salt, the solubility of glycine decreases greatly with decreasing D' . This effect has already been discussed above. (2) The addition of salt increases the solubility of glycine, and the rate of increase, for a given increment in ionic strength, is greater, the lower the dielectric constant. Increase in solubility is of course equivalent to a decrease in the activity coefficient, f_A (equation 94). This is the type of ion-dipole interaction that is envisaged in Kirkwood's theory. In any given solvent, we wish to determine the value of the coefficient K , which is the change in the logarithm of the *relative* activity coefficient of the solute with changing ionic strength in a given solvent. If f_A° is the activity coefficient in the given solvent at zero ionic strength, referred to dilute aqueous solution as the standard medium in which $f_A = 1$ by definition (see Fig. 15), and f_A is the activity coefficient of the solute in the same medium but at finite ionic strength (ω), then the relative activity coefficient (γ) is defined by

$$\log \gamma = \log (f_A/f_A^\circ) = \log (N_A^\circ/N_A) \quad (96)$$

$\log \gamma$, as here defined, corresponds to $\log f$ in (87), (88), (89), and (90). Since the factor K in (88), (89), and (90) is inversely proportional to the square of the dielectric constant (if the salting-out term in equation 88 is neglected), it is convenient to plot all the data in different media on a comparable basis by multiplying $\log \gamma$ by D/D_0 , and to represent this as a function of $(D_0/D)\omega$. If the salting-out term in (88) can be neglected, and if the theory is correct in predicting that the salting-in term, K , is proportional to $1/D^2$, then a plot of the data in several media of different dielectric constants should bring all the curves into superposition, for a given amino acid and a given salt. The results of this procedure for the data in 60%, 80%, 90%, and 95% ethanol are shown in Fig. 16.

This figure presents the same data represented in the four lower curves of Fig. 15 for glycine in lithium chloride solutions. All the curves now start from a common origin, however, since the ordinate in each case is proportional to $\log \gamma$, which from equation (96) necessarily becomes equal to zero at zero ionic strength in any given medium. It is seen that the four sets of points, for the four different ethanol-water mixtures, with D' values for the pure solvents ranging from 47.7 to 27.1, all fall on a common curve when plotted in this way.¹⁶ This is strong confirmation of the theo-

¹⁶ It should be noted that the value of D in Fig. 16 is not the dielectric constant (D') of the pure solvent but is the dielectric constant of the medium, taking account of the concentration of glycine present at saturation. Glycine, like other dipolar ions,

retical prediction that the salting-in term of the coefficient K in (88) is inversely proportional to D^2 . In media of relatively low dielectric constant, such as those under discussion here, the salting-out term in (88), which is proportional to b^3/a , is relatively unimportant, since it varies inversely as the first power of D , and therefore becomes less and less important, relative to the salting-in term, as D decreases.

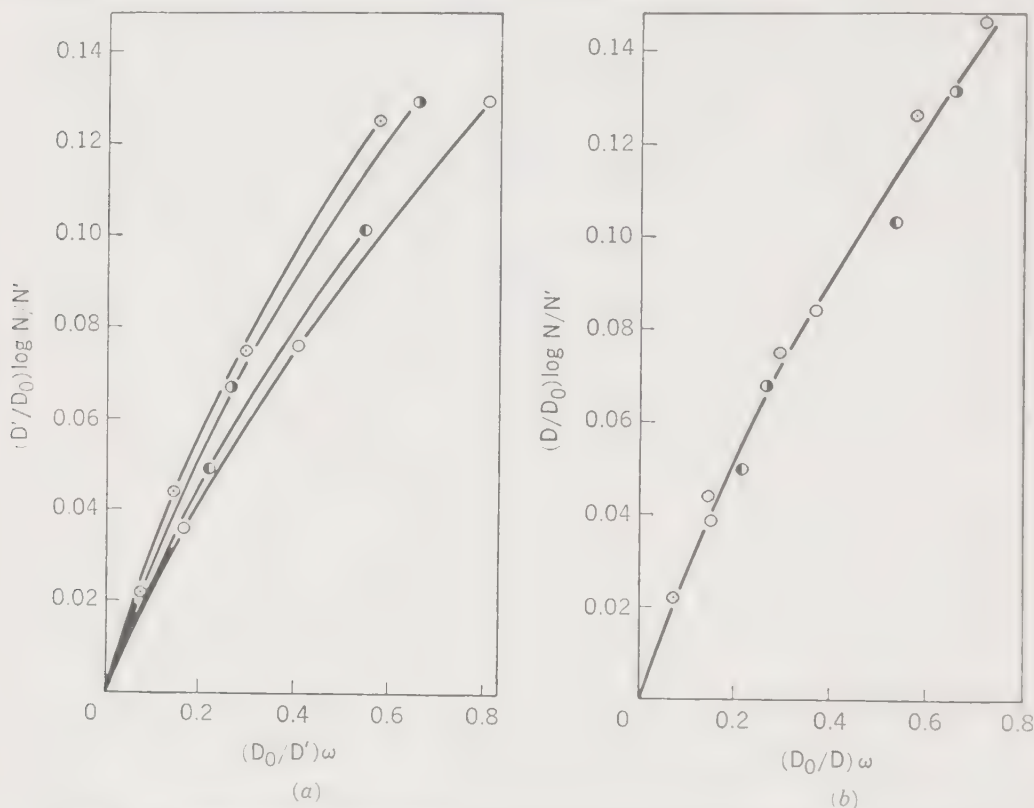


FIG. 16. Interaction of glycine and lithium chloride in ethanol-water mixtures at 25°. In the left-hand diagram the data are plotted as a function of the dielectric constant D' of the pure salt-free solvent in the absence of dissolved glycine. In the right-hand diagram the data are plotted in terms of D , the dielectric constant of the saturated solution, including the effect of the dissolved glycine. (From Cohn and Edsall, 1943, p. 272.)

We may now use the experimental data on glycine in salt solutions, summarized in Figs. 15 and 16, to calculate the dipole moment of this dipolar ion, considering it as a sphere of radius b , the mean distance of closest approach between the dipolar ion and the surrounding ions being

produces a large positive increment in the dielectric constant of the medium, increasing D by approximately 22.5 units per mole of glycine per liter. The magnitude of this dielectric increment is a function of the dipole moment of the dipolar ion. The theory of this relation will be taken up in detail in Chapter 6. In view of the magnitude of this dielectric increment of glycine the correction is important.

a (Fig. 17). The positive charge, $+\epsilon$, and the negative charge, $-\epsilon$, are separated by the dipole distance, R . If R is in centimeters and ϵ in electrostatic units, the dipole moment $\mu = 4.8 \times 10^{-19}R$. If we express R in angstroms, and μ in Debye units (1 Debye unit = 10^{-18} esu), then $\mu = 4.8R$. If we take water as our standard reference medium ($D_0 = 78.54$ at 25°) and substitute numerical values for the constants in (88), then the

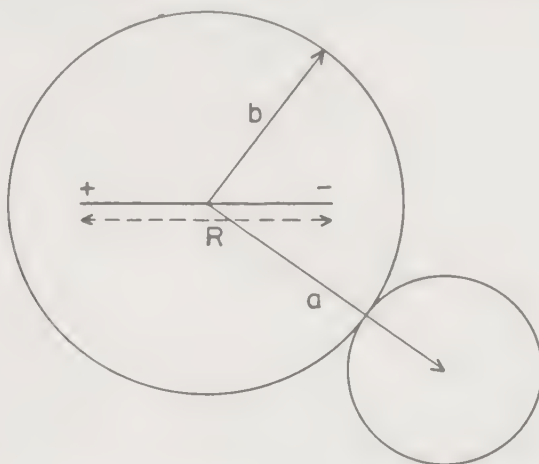


FIG. 17. Model of a spherical dipolar ion containing a positive charge $+\epsilon$ and a negative charge $-\epsilon$ separated by the dipolar distance R .

salting-in term in (88) may be written, if R is in angstroms and μ in Debye units:

$$\begin{aligned} K_R &= 0.00548(D_0/D)^2(\mu^2/a) \\ &= 0.125(D_0/D)^2(R^2/a) \end{aligned} \quad (97)$$

To correlate measurements in media of different dielectric constant, such as those in Figs. 15 and 16, it is convenient to define a coefficient, K_R' , equal to $(D/D_0)^2 K_R$:

$$K_R' = \lim_{\omega \rightarrow 0} \frac{d(D/D_0) \log_{10} (N_A/N_A^\circ)}{d(D_0/D)\omega} = 0.00548\mu^2/a \quad (98)$$

This is valid if the salting-out term in (88) can be neglected. Experimentally K_R' , for glycine in lithium chloride solutions, is found from the limiting slope of the curves in Fig. 16 to be 0.32. From data on the partial molal volume of glycine and related substances, E. J. Cohn estimated its radius, b , to be 2.80 Å, and estimated a for glycine in lithium chloride solutions as 3.90 Å. Substituting these values in (98), we find $\mu = 15.1$ Debye units. From a space model, based on interatomic distances and bond angles in amino acid crystals (Corey, 1948, p. 391), we estimate a distance very close to 3.0 Å between the positively charged nitrogen and the center of the line joining the two oxygens of the $-\text{COO}^-$ group, which

may be taken as the center of negative charge. This gives a dipole moment $\mu = 3.0 \times 4.8 = 14.4$ Debye units, in close agreement with that obtained from salting in. A nearly identical value is obtained from analysis of the dielectric increments of glycine in solution, as we shall see in Chapter 6.

In 60% ethanol, and in media of lower dielectric constant, the analysis just given shows that, as a first approximation, we are justified in ignoring the salting-out term in (88). In aqueous solution, however, this is not true. The solubility data for glycine in water and aqueous salt solutions cannot be used to calculate values of the interaction coefficient, K ; the solubility is so great, even in the absence of salt, that the interactions between the glycine dipoles obscure the effect of the ion-dipole interactions. Values of K , however, for aqueous solutions of salts in glycine, have been obtained by Joseph (1935) from electromotive force measurements, and by Scatchard and Prentiss (1934) from freezing point measurements. This gives $K = 0.24$, for glycine in sodium chloride solutions. The difference between this value and K_R' in (98), which is 0.32, gives a salting-out term very nearly 0.08. Since $b = 2.80$ Å, and $a = 4.05$ Å approximately for glycine and sodium chloride, this gives $\rho = b/a = 0.69$, and the function $\alpha(\rho) = 1.37$ (Kirkwood, 1943). These numerical values give for the second term in (88) a value of -0.08 , in excellent agreement with the value inferred from experiment as described above. Thus the salting out due to image forces is of the order of 25% of the salting-in effect for glycine.

The contrast between the relatively large salting-out effect in aqueous solutions and the predominant salting-in effect in media of low dielectric constant is even more strikingly shown in α -aminocaproic acid or leucine, which contain large hydrocarbon side chains. Leucine, when dissolved in water, is actually salted out by sodium chloride or potassium chloride (see, for instance, the figure on p. 238 of Cohn and Edsall, 1943). In 90% ethanol, however, it is salted in to almost the same extent as glycine, and the same is true of α -aminocaproic acid (see Fig. 18).

It is illuminating to consider the magnitude of the salting-in effect for a series of dipolar ions of different dipole moment. Data for a group of such compounds, in 80% ethanol containing sodium chloride and in 95% ethanol containing lithium chloride, are shown in Fig. 18. The data in this figure are plotted in just the same manner as in Fig. 16, and it will be noted that the salting-in effects for the dipoles of higher electric moment, as characterized by the salting-in constant, K_R' , are much greater than for glycine. In comparing the K_R' values with the dipole moments of these dipolar ions, we make use of the data given in Chapter 6, in which values of μ are calculated from dielectric constant measure-

ments on solutions of dipolar ions. The dipole moments so derived are compared in Table III with the salting-in constants given by the limiting slopes of the curves in Fig. 18. It will be seen that there is a direct proportionality between the salting-in constant and the estimated dipole moment, which holds closely over a fourfold range of variation in both

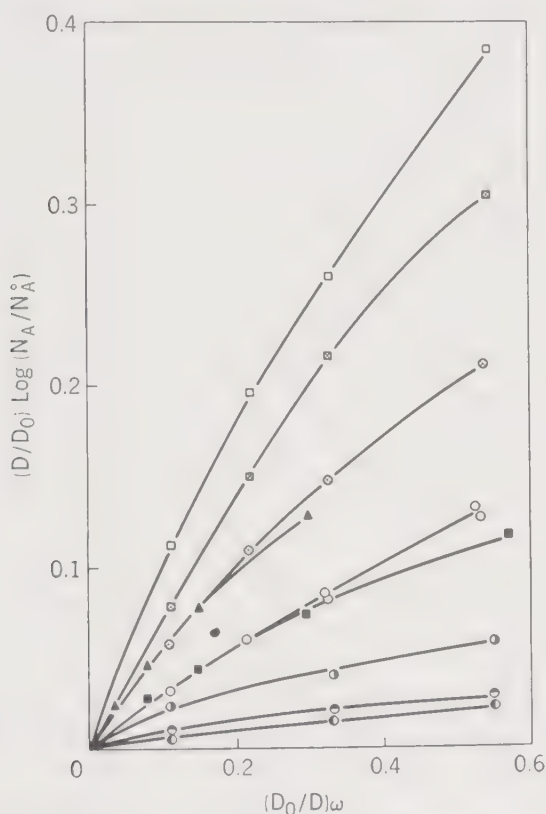


FIG. 18. Relative solubility of amino acids and peptides in 80% ethanol containing sodium chloride. Reading from top to bottom, the curves are for: lysylglutamic acid; triglycine; diglycine; ϵ -aminocaproic acid (black triangles); glycine; α -aminocaproic acid (black squares); hydantoic acid of diglycine; hydantoic acid; hydantoic acid of α -aminocaproic acid. The measurements on α - and ϵ -aminocaproic acids were carried out in 95% ethanol containing lithium chloride. (From Cohn and Edsall, 1943, p. 274.)

K_R' and μ . This proportionality between K_R' and the *first* power of the dipole moment would be predicted from Kirkwood's first ellipsoidal model (equation 89). This is to be contrasted with the spherical model, for which it is predicted that K_R' should be proportional to the *square* of the dipole moment. We shall not attempt to discuss further the use of Kirkwood's equations (89) and (90); the interested reader should refer to the account given by Kirkwood himself (in Cohn and Edsall, 1943, Chapter 12).

The chemist who is interested in the use of salting-in and salting-out procedures for purposes of fractionation is naturally concerned with the solubility of proteins and other substances over a wide range of salt concentrations, far beyond the narrow region in which Kirkwood's limiting laws can be expected to hold. An example of the data obtained, when studies are carried out over a wide range of salt concentrations, is shown in Fig. 19. This shows the solubility of cystine in four different salts. It will be noted that the salting-in effects at low ionic strength differ markedly for the different salts, being greatest for calcium chloride and least for sodium sulfate. Cystine is a considerably larger molecule than glycine, and it contains two positively and two negatively charged groups. Therefore a somewhat more complicated model would be required to

TABLE III
SALTING-IN CONSTANTS, K_R' , AND DIPOLE MOMENTS OF CERTAIN AMINO ACIDS AND PEPTIDES

Substance	Salting-in constant, K_R'	Dipole moment, μ	μ/K_R'
Glycine	0.33	15.5	47
α -Aminocaproic acid	0.33	(15.5)	(47)
Diglycine	0.58	27.6	48
ϵ -Aminocaproic acid	0.7	29	41
Triglycine	0.8	35	44
Lysylglutamic acid	1.2	61	51

The dipole moment of α -aminocaproic acid has been assumed to be the same as for glycine, since both are α -amino acids. See the discussion of dipole moments of amino acids in Chapter 6.

represent it than the simple dipole, enclosed within a sphere, which we used for glycine. Nevertheless, it is reasonable to assume that a limiting equation similar in general character to (88) would be suitable to describe the salting in (see Kirkwood, 1934). Clearly a salting-out term must be included in the limiting equation also, and the magnitude of this term must increase for the different salts in the order $\text{CaCl}_2 < \text{NaCl} < (\text{NH}_4)_2\text{SO}_4 < \text{Na}_2\text{SO}_4$.

We note here a fundamental difference between the limiting laws for ion-ion and ion-dipole interactions. Since $-\log f_{\pm}$, for ion-ion interactions, is proportional to $\sqrt{\omega}$ at very low ionic strengths (equation 76), whereas the salting-out effect is proportional to ω (equation 51), it follows that only the term in $\sqrt{\omega}$ remains in the limiting law, which thus is a function of the total ionic strength only, and does not depend on specific characteristics of individual ions. For ion-dipole interactions, however,

both the salting-in and the salting-out terms are proportional to ω , and the salting-out term therefore does not vanish relative to the other, as ω approaches zero. Hence the limiting slope, K , depends not only on the dimensions and dipole moment of the dipolar ion, but also on the dimensions of the salt ions, which enter into the parameter a in (88). It may be noted also that in solutions of very high ionic strength (above 6), the salting-out of cystine in ammonium sulfate solutions shows the simple linear relation between the logarithm of the solubility and the ionic strength, which we have already discussed. The salting-out constant for

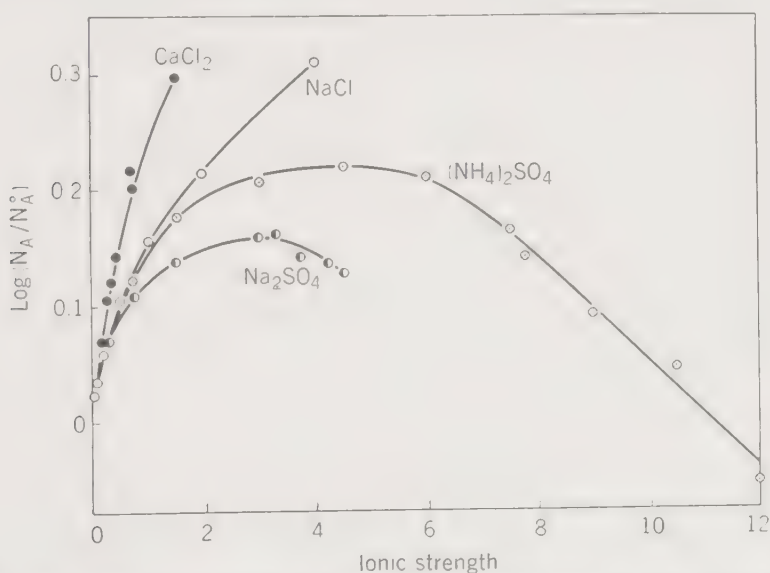


FIG. 19. Solubility of cystine in aqueous salt solutions. (From Cohn and Edsall, 1943, p. 242.)

cystine in ammonium sulfate is included in Table I. The salting-out effect observed here, however, at high salt concentrations, must evidently be explained otherwise than the salting-out term which enters into the expression of the limiting law in (88). For there is nothing in the limiting law to explain why the two lower curves in Fig. 19, after rising with increasing ω at low ω values, should reach a maximum and then descend rather steeply as ω increases still further. Other factors, relatively insignificant at low ionic strength, must come into play at very high ionic strength. The principal factor, as has been suggested before (p. 278), may be essentially a dehydration effect; each ion binds some water of hydration around it by very strong electrostatic forces. When a great many ions are present, so large a fraction of the total water may be bound in this fashion that the amount available to act as a solvent for the cys-

tine is considerably reduced. As yet, however, no quantitative treatment of the salting-out effect along these lines has been given.

Qualitatively there is striking similarity between these data for cystine and the data of A. A. Green for the solubility of horse carboxyhemoglobin in salt solutions of varying ionic strengths, which are represented in Fig. 20. It is plain, on comparing Figs. 19 and 20, that sodium sulfate is a more effective salting-out agent than ammonium sulfate, whereas sodium chloride has a much weaker salting-out action than either, for both the amino acid and the protein. The scale of Figs. 19 and 20 is very different, however. The salting out of carboxyhemoglobin in

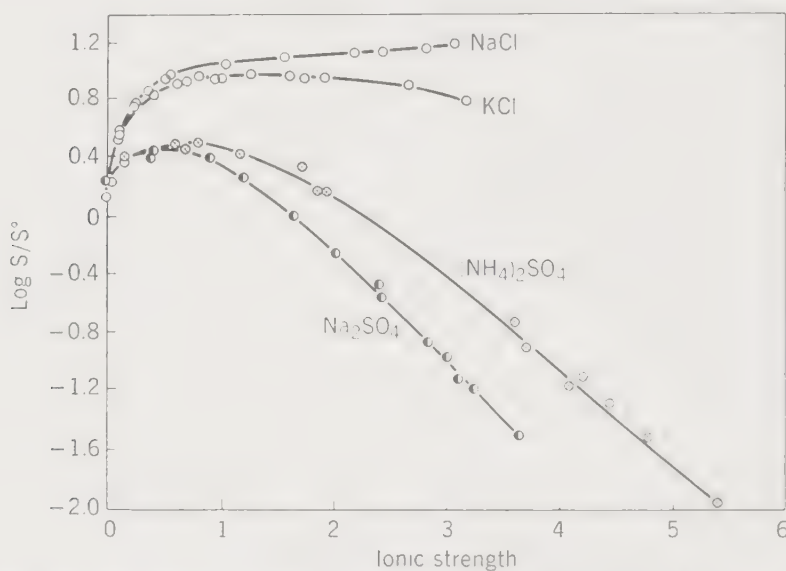


FIG. 20. The solubility of horse carboxyhemoglobin in salt solutions of varying ionic strength. Note the difference in scale of both abscissa and ordinate as compared with Fig. 19. (Data of A. A. Green. From Cohn and Edsall, 1943, p. 608.)

ammonium sulfate has already begun at ionic strength below 2, and the solubility falls by a factor of nearly 100 between ionic strengths 2 and 5. In contrast, the solubility of cystine decreases by a factor of less than 2 between ionic strength 6 and 12. In other words, the interactions of cystine with salts are qualitatively very similar to those of hemoglobin, but quantitatively the effects of salt on the protein are enormously greater. Indeed, the salting out of proteins is so striking an effect, and so easily observed, that it has been known for at least a century, whereas the similar effects for cystine were found only after 1930. It is interesting to note, however, that in principle the same types of interactions occur between the very large protein molecules and salt ions as those which can

be detected between the same salts and a simple amino acid, such as cystine.

The magnitude of the salting-in effects differs greatly from one type of dipolar ion to another. Some examples are illustrated for two proteins and several amino acids in Fig. 21, which shows the action of sodium chloride on the solubility of β -lactoglobulin, horse carboxyhemoglobin, cystine, asparagine, and glycine. The salting-in effect for β -lactoglobulin is enormous, the solubility increasing about one hundredfold when the

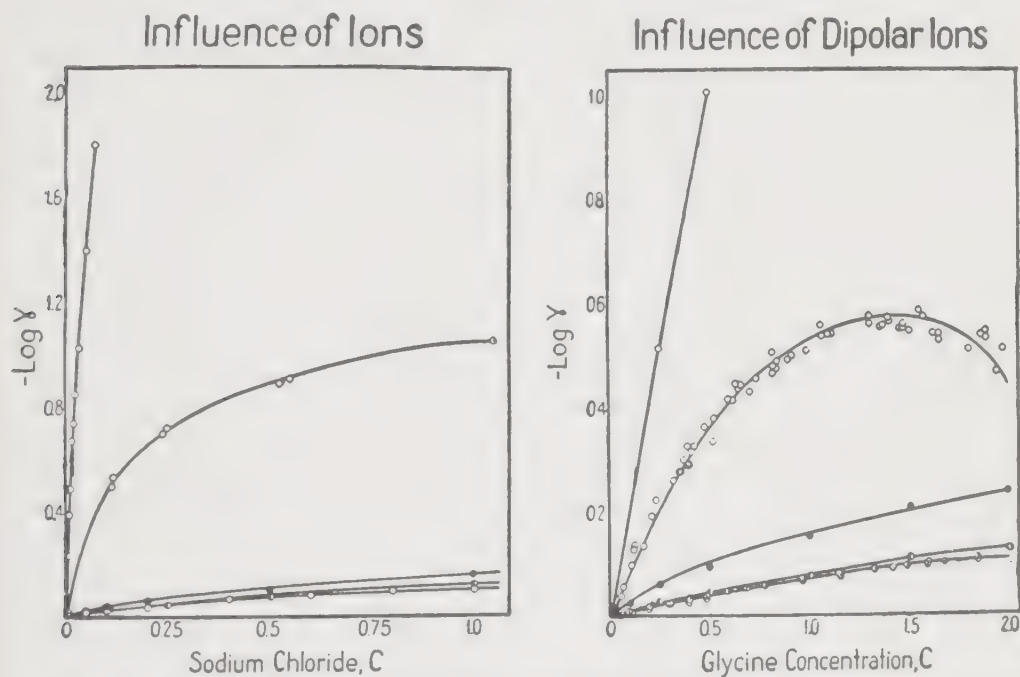


FIG. 21. Relative effects of sodium chloride and glycine on the solubility of certain proteins and amino acids. Note differences in scale of both abscissa and ordinate in the two diagrams. In each figure, the steepest curve is for β -lactoglobulin, then below in descending order: horse carboxyhemoglobin, cystine, asparagine, glycine. (From Cohn and Edsall, 1943, p. 420.)

ionic strength increases from 0 to 0.1. The effect on hemoglobin is much less, and that on the three amino acids is of an altogether lower magnitude. It is interesting and important to note that qualitatively similar "salting-in" effects are produced by the action of one dipolar ion on another. This is illustrated in the right-hand diagram of Fig. 21, which shows the effect of glycine in increasing the solubility (or decreasing the activity coefficient) of the same series of molecules for which the effect of sodium chloride is shown on the left-hand diagram. Qualitatively, the order of the effects is the same for both reagents, the solubility of β -lactoglobulin being far more affected than any of the others by either glycine

or sodium chloride. It will be observed, however, that glycine, at a given molar concentration, is a far less effective agent for dissolving these proteins than sodium chloride (note carefully the different scales for abscissa and ordinate in the two parts of Fig. 21). Actually the parallelism breaks down completely in certain cases; for example, edestin, the globulin of hemp seed, is very readily salted in by sodium chloride, but glycine appears to have no salting-in action on it at all. These effects of glycine on the solubility of other dipolar ions serve, however, to illustrate some of the phenomena of dipole-dipole interactions. These phenomena have been discussed further elsewhere (for reference see, for instance, Cohn and Edsall, 1943, Chapter 10; Gucker, Klotz, and Allen, 1942).

In pointing out these analogies between the behavior of amino acids and that of proteins we should note one distinction of great importance. In solubility studies on amino acids it is almost always possible to show that the composition of the solid phase is that of the pure amino acid (or in some cases of one of its hydrates) and that this solid phase remains the same when it is equilibrated with salt solutions over a wide range of composition. This is by no means true of proteins, however. Protein crystals contain large quantities of water, of the order of 50% by weight, and sometimes more. Moreover, if salt is present in the mother liquor surrounding the protein crystals it also penetrates into the crystals themselves, and the salt and water composition of the crystals varies with that of the medium. A detailed study of these phenomena for β -lactoglobulin was made by McMeekin and Warner (1942). Extensive further studies on crystals of this protein, as well as on serum albumin and insulin, were carried out by Low and Richards (1954). (See also Lewin, 1951, for serum albumin.)

In view of this variable composition of protein crystals it is clearly not justifiable to assume without further evidence that the chemical potential of the protein component in the crystal is constant, and independent of the variations in the composition of the crystal. In other words, the relative solubility of a given protein, in two media of different ionic strengths, cannot in general be taken as giving directly the ratio of the activity coefficients in the two media, as we have assumed, with strong justification, for the case of amino acids (see equations 95 and 96). Nevertheless, effects of ionic strength on solubility are qualitatively so similar that there seems little doubt that the same general types of interactions are involved in both cases.

GENERAL REFERENCES

There are numerous excellent treatments of the general principles of electrostatics. Here we name only a few. As a preliminary remark we strongly recommend to the

reader who wishes to go further that he master the principles of vector analysis, if he has not already done so, before proceeding to a deeper study of the subject. The fundamental relations of vector analysis are readily learned, and their use greatly simplifies the development of the subject. A good brief treatment of vector analysis is given, for instance, in the introductory chapter of L. Page (1935), "Introduction to Theoretical Physics," 2nd ed., D. Van Nostrand and Co., Princeton, New Jersey. The chapters on electricity and magnetism in the same book include an excellent short treatment of electrostatics. J. H. Jeans (1927), "The Mathematical Theory of Electricity and Magnetism," Cambridge University Press, gives a very extensive treatment of classical electrostatics with an extensive mathematical development built up from the fundamentals. Jeans, however, does not make use of vector analysis in his presentation.

Two more recent texts of high quality are:

- Peck, E. R. (1953). "Electricity and Magnetism," McGraw Hill Book Co., New York, Chapters 1-4 inclusive.
Panofsky, W. K. H., and Phillips, M. (1955). "Classical Electricity and Magnetism," Addison-Wesley Publishing Co., Reading, Massachusetts, Chapters 1-6 inclusive.

For many of the fundamental mathematical relations, the treatment of H. Margenau and G. M. Murphy (1955), "The Mathematics of Physics and Chemistry," 2nd ed., D. Van Nostrand and Co., Princeton, New Jersey, may be recommended.

The papers of Debye and his collaborators on the salting-out phenomenon and on interionic attraction theory have been reprinted in English translation with many other important contributions in "The Collected Papers of Peter J. W. Debye," (1954), Interscience Publishers, New York. Specific references to some of these papers are given below, under Special References.

A good, relatively simple discussion of the Debye-Hückel theory is given by D. A. MacInnes (1939), "The Principles of Electrochemistry," Reinhold Publishing Corp., New York. A discussion at a somewhat more advanced level is given by G. Scatchard in Chapter 3 of "Proteins, Amino Acids and Peptides," by E. J. Cohn and J. T. Edsall (see references below). A critical discussion of the theory, at a highly advanced level, is given by R. H. Fowler and E. A. Guggenheim (1952), "Statistical Thermodynamics," Cambridge University Press, Chapter IX.

SPECIAL REFERENCES

- Cohn, E. J. (1925). *Physiol. Revs.* **5**, 349.
Cohn, E. J., and Edsall, J. T. (1943). "Proteins, Amino Acids and Peptides," Reinhold Publishing Corp., New York.
Corey, R. B. (1948). *Advances in Protein Chem.* **4**, 385.
Debye, P. (1927). *Z. physik. Chem.* Cohen Festband, p. 56; also in "Collected Papers of P. J. W. Debye," p. 366, Interscience Publishers, New York, 1954.
Debye, P., and Hückel, E. (1923). *Physikal. Z.* **24**, 185, 305; also in "Collected Papers of P. J. W. Debye," p. 217, 264, Interscience Publishers, New York, 1954.
Debye, P., and McAulay, J. (1924). *Physikal. Z.* **26**, 22. Also in "Collected Papers of P. J. W. Debye," p. 355, Interscience Publishers, New York, 1954.
Edsall, J. T. (1947). *Advances in Protein Chem.* **3**, 383.
England, A., Jr., and Cohn, E. J. (1935). *J. Am. Chem. Soc.* **57**, 634.
Glasstone, S., and Pound, A. (1925). *J. Chem. Soc.* **127**, 2665.

- Green, A. A. (1931). *J. Biol. Chem.* **93**, 495, 517.
- Gross, Ph., and Iser, M. (1930). *Monatsh. Chem.* **55**, 329.
- Gross, Ph., and Schwarz, K. (1930). *Monatsh. Chem.* **55**, 287.
- Gucker, F. T., Jr., Klotz, I. M., and Allen, T. W. (1942). *Chem. Revs.* **30**, 191.
- Harned, H. S., and Owen, B. B. (1950). "The Physical Chemistry of Electrolytic Solutions" 2nd ed. Reinhold, New York.
- Joseph, N. R. (1935). *J. Biol. Chem.* **111**, 479, 489.
- Kirkwood, J. G. (1934). *J. Chem. Phys.* **2**, 351.
- Kirkwood, J. G. (1943). In "Proteins, Amino Acids and Peptides," by E. J. Cohn and J. T. Edsall, Chapter 12, Reinhold Publishing Corp., New York.
- Lewin, J. (1951). *J. Am. Chem. Soc.* **73**, 3906.
- Linderstrøm-Lang, K. (1953). *Compt. rend. trav. lab. Carlsberg Sér. chim.* **28**, 281.
- Long, F. A., and McDevit, W. F. (1952). Activity coefficients of nonelectrolyte solutes in aqueous salt solutions. *Chem. Revs.* **51**, 119.
- Low, B. W., and Richards, F. M. (1954). *J. Am. Chem. Soc.* **76**, 2511.
- McMeekin, T. L., and Warner, R. C. (1942). *J. Am. Chem. Soc.* **64**, 2393.
- Randall, M., and Failey, C. F. (1927). *Chem. Revs.* **4**, 271, 285.
- Robinson, R. A., and Stokes, R. H. (1955). "Electrolyte Solutions," Academic Press, New York, and Butterworth's Scientific Publications, London.
- Scatchard, G. (1927). *Chem. Revs.* **3**, 383.
- Scatchard, G. (1941). *J. Chem. Phys.* **9**, 34.
- Scatchard, G., and Benedict, M. A. (1936). *J. Am. Chem. Soc.* **58**, 837.
- Scatchard, G., and Prentiss, S. S. (1934). *J. Am. Chem. Soc.* **56**, 2314.
- Smith, P. K., and Smith, E. R. B. (1937). *J. Biol. Chem.* **117**, 209; **121**, 607.
- Smith, P. K., and Smith, E. R. B. (1940). *J. Biol. Chem.* **132**, 47, 57.

Chapter 6

Dielectric Constants and Their Significance

Introduction to Dielectric Constants and Dipole Moments

In the presentation of the basic ideas of electrostatics given in Chapter 5, we started with the concept of point charges embedded in a dielectric medium which was treated as if it were continuous, without internal structure, and a perfect insulator. The charges in this hypothetical medium were assumed to attract or repel one another in accordance with Coulomb's law, that is, the inverse square law, with the introduction of an empirical dimensionless constant, the dielectric constant (D), characteristic of each insulator. No attempt was made to explain the variability of D from one medium to another or to relate it to molecular properties. We also introduced, without attempting any sharp definition, the concept of a conductor, and in dealing with the free energy of ions in solution, we adopted the model of conducting spheres embedded in an insulator.¹

¹ It is perhaps worth pointing out that there are various other ways in which we might have approached the subject. Each of these simply represents a particular preference as regards the way in which we formulate the experimental picture. Each has its particular advantages and disadvantages. For example, instead of beginning with the idea of electric charge and introducing Coulomb's law, we might have started with the concept of an electromagnetic field. The fundamental ideas would then have been the electric and magnetic force vectors which characterize the field, and which determine respectively the electric and magnetic energies of the field and can be defined in terms of these with the introduction of two arbitrary constants, the dielectric constant and the magnetic permeability. Instead of Coulomb's law, we might then have introduced, as the embodiment of the experimental facts, the law known as Poynting's law, which relates the flux of energy in the field to these two vectors. From this, with the aid of the law of the conservation of energy, we could then have derived at once Maxwell's equations, as the governing equations of electromagnetic theory. Gauss's law, and the associated concept of electric charge, both surface charge and space charge, would have followed naturally. By setting the time derivatives in Maxwell's equations equal to zero, corresponding to the static case, the electric force would have then shown itself to be derivable from a potential, and the principles of classical electrostatics would have resulted without further resort to experimental facts. In this procedure only one other constant is involved in addition to the dielectric constant and the magnetic permeability, namely a constant occurring in Poynting's law. By making the dielectric constant and magnetic permeability dimensionless, as in the familiar Gaussian system of units, this other constant turns out to have the dimensions of a velocity, and in the development of the theory can be

In classifying all bodies as either insulators or conductors, we introduce an idealization, though a useful one, and in treating the insulator as a continuum, we close the door on an interpretation of the dielectric constant in terms of molecular properties, a subject which has thrown much light on many aspects of physical chemistry in recent years and which will be the theme of this chapter. In order to lead up to it, it will be well to reconsider our concepts somewhat. In doing so we may appropriately begin with a few words about insulators and conductors.

In a hypothetical perfect conductor, a charged particle would move freely, without resistance, in the presence of an electric field, and no heat would be generated by the motion, since the resistance would be zero. Many metals at extremely low temperatures become "superconductive" and approach the perfect conductor in the ease with which they permit electrons to flow through them. At higher temperatures they are still good conductors, but decidedly imperfect, displaying a readily measurable resistance to the flow of current. A perfect insulator, by contrast, is a hypothetical medium in which the resistance is infinite, so that no current can flow at all. Most ordinary gases can be treated generally as perfect insulators, provided they do not contain ionic particles and are not subjected to extremely high electric fields. Actually any real medium, in an electric field, is always the seat of some energy dissipation in which electric energy is transformed into heat. Nevertheless in certain media such energy dissipation proceeds so slowly, in terms of the time scale of the processes with which we are concerned, that it may be neglected and the field regarded as static. In classical electrostatics, therefore, where we are not concerned with the flow of current, it is customary to make a somewhat arbitrary separation and distinguish only between conductors and insulators.

We now turn to the concept of surface charge. Since, if there is no flow of electric charges, there can be no field in the interior of a conductor, it follows from Gauss's law (Chapter 5, Equations 6 and 8) that there can be no space charge at any point within it. All the charges must collect, therefore, at the surface, and give rise to a sheet or layer of surface charge. Such a sheet of charge is characterized at any point by its surface den-

shown to be identifiable with the velocity of light. It is, however, also possible, as is actually done in the so-called Maxwellian electrostatic system of units, to make the constant of Poynting's law dimensionless. Then, by keeping the dielectric constant as a dimensionless quantity, the magnetic permeability acquires the dimensions of the reciprocal of the square of a velocity. If, alternatively, as in the Maxwellian electromagnetic system of units, we keep the magnetic permeability as dimensionless, the dielectric constant assumes these dimensions, which illustrates the arbitrariness of conventions regarding the dimensions of physical constants.

sity, which may be calculated by Gauss's law. If as a closed surface, we choose an infinitesimal squat cylinder, its faces parallel to an element of surface of the conductor, one just inside and one just outside the conductor, and make the sides of the cylinder negligible in area in relation to

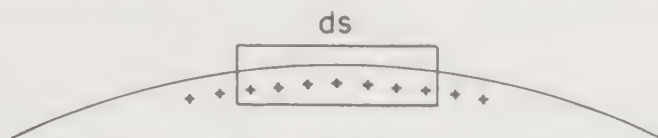


FIG. 1. The electric field due to a surface density of charge.

the faces, it follows that the charge dq on the element of surface ds of the conductor enclosed by the cylinder is given (Fig. 1) by

$$4\pi dq = dN = DE_n ds$$

where E_n is the component of the electric intensity just outside the conductor and normal to its surface, positive in the direction pointing towards the insulator, and D is the dielectric constant of the insulator. Consequently, the surface density of charge $\sigma = dq/ds$ is given by

$$\sigma = \frac{DE_n}{4\pi} \quad (1)$$

In the classical theory of electrostatics, in view of the finite conductivity of all actual substances, it is usual to extend the principle of the vanishing of the space charge inside a conductor to all media, quite generally, with the result that all charges, except for certain point charges or point dipoles introduced to meet special conditions, are supposed to exist as surface charges, distributed over the interfaces between insulators and conductors, and, more rarely, the interfaces between insulators. In any medium, therefore, Laplace's equation for the potential ($\nabla^2\psi = 0$) is supposed to be everywhere applicable, though solutions of the equation in different regions separated by a surface of discontinuity are expected to be different. The introduction of the point charges or dipoles does not affect this, though it does give rise to the existence of point singularities—points where the potential becomes infinite—one for each such charge or dipole.² Except for such singularities the potential is everywhere continuous, for passing through a sheet of charge of finite density does not involve a discontinuity in the potential.

² In the Debye-Hückel theory of ion-ion interactions, there is a departure from these principles in the assumption of a space charge in the ion atmosphere surrounding the ion under consideration and the use of Poisson's equations to calculate the potential.

Just as in the case of the charge at the surface of a conductor, so too the charge at the interface between two insulators can be calculated by Gauss's law, making use of an infinitesimal squat cylinder enclosing an element at the interface. Here, however, it is necessary to take account of the electric intensity in *both* media. The result is

$$4\pi\sigma = D_1E_{n_1} + D_2E_{n_2} \quad (2)$$

in which the subscripts 1 and 2 serve to distinguish between two insulators. It should be emphasized that E_{n_1} and E_{n_2} are the *inward* normal components of the electric intensity in each medium; if E_n points outward in either medium it is taken as negative in that medium. In the case of two insulators, there may be tangential components of E at the interface (it can be shown that they must be equal on the two sides of it), but for an insulator in contact with a conductor, this is prevented by the vanishing of the field in the conductor. In this case the electric vector is, therefore, everywhere perpendicular to the surface.

Equations (1) and (2) are somewhat simplified if we make use of the quantity known as the electric displacement, \mathbf{D} . This is defined for any medium as

$$\mathbf{D} = D\mathbf{E} \quad (3)$$

Like \mathbf{E} , \mathbf{D} is a vector, since it is the product of a vector, \mathbf{E} , with the simple scalar quantity D . In terms of the normal components of \mathbf{D} , denoted by D_n , equations (1) and (2) become respectively

$$D_n = 4\pi\sigma \quad (4)$$

$$D_{n_1} + D_{n_2} = 4\pi\sigma \quad (5)$$

When there is no charge ($\sigma = 0$) at the interface between the two insulators, the normal component of the electric displacement is a continuous function across the boundary.

The relation between \mathbf{D} and \mathbf{E} , in terms of the dielectric constant of the medium, is given by (3). Another very important relation between them may be simply derived by considering a parallel plate condenser, with a uniform surface density of charge equal to σ at the positive plate, and to $-\sigma$ at the negative plate. We assume the plates to be large, so that the electric intensity, E , between them is uniform, except in small regions near the edges of the plates, which we shall neglect. If there is a vacuum between the plates, the electric intensity is normal to the plates and equal to

$$E_0 = 4\pi\sigma$$

If a slab of dielectric is now introduced between the plates without any change in the actual surface charge, σ , the electric intensity falls from

E_0 to E_0/D . The effect is the same as if, with a vacuum maintained between the plates, the surface charge had been reduced from σ to σ/D . Thus the dielectric medium between the plates may be considered as contributing a charge (of opposite sign):

$$\sigma' = \sigma - \frac{\sigma}{D} = \left(\frac{D-1}{D} \right) \sigma = \frac{D-1}{4\pi} E \quad (6)$$

to each of the condenser plates per unit area. This is negative on the surface in contact with the positive plate. These are induced charges, due to the polarization of the molecules of the dielectric medium by the electric field of the condenser.

Let d be the distance between the plates (Fig. 2). Consider an element of volume in the dielectric between the plates—say a uniform small cylinder of cross section A , with its axis perpendicular to the plane of the plates. The charge on the left-hand face of the cylinder is $-\sigma'A$; that on the right-hand face is $+\sigma'A$; and the charges are separated by the distance d . Thus the cylinder is a dipole of moment $\sigma'Ad$. Like any dipole it is a vector; the direction of the vector is taken by convention as pointing from the negative to the positive end of the dipole, that is, from left to right in Fig. 2. The volume of the cylinder is Ad ; hence σ' is the electric moment per unit volume. Obviously any other element of volume in the dielectric between the plates is characterized by the same electric moment per unit volume, which is generally denoted by the symbol \mathbf{I} . In an isotropic medium, \mathbf{I} is a vector parallel to \mathbf{E} .

Since $\mathbf{I} = \sigma'$, we can immediately rewrite (6) in the form

$$(D-1)\mathbf{E} = 4\pi\mathbf{I} \quad (7)$$

The quantity $(D-1)/4\pi$, which determines the magnitude of \mathbf{I} in relation to \mathbf{E} , is known as the electric susceptibility. Equation (7) is one of the fundamental equations of electrostatics. It is often rewritten in terms of the electric displacement in the form

$$\mathbf{D} = \mathbf{E} + 4\pi\mathbf{I} \quad (8)$$

If, therefore, we consider the charge due to polarization on any free sur-

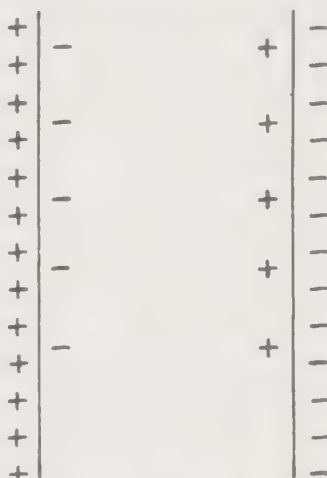


FIG. 2. Simplified diagram of a parallel plate condenser with a dielectric medium between the plates.

face of the dielectric whose normal makes an angle θ with \mathbf{E} , the density of charge on this surface is

$$\sigma_{\theta}' = I \cos \theta = \frac{D - 1}{4\pi} E \cos \theta \quad (9)$$

It is of interest to observe that the charge due to polarization is the same as the image charge in the second example discussed in Chapter 5, p. 261, namely, the image charge at the spherical interface between two dielectrics produced by a point charge at the center of the sphere.

Dielectric Polarization in Relation to Molecular Properties

THE DEBYE THEORY OF THE DIELECTRIC CONSTANT

The question with which we are confronted is how an actual discontinuous medium, made up of molecules, can be supposed to become polarized, that is, develop an electric moment, under the influence of an applied electric field. Two possibilities present themselves. According to the first, we think of the molecules as aggregations of charges—positive charges of atomic nuclei and negative charges of electron shells—which suffer a displacement under the influence of an applied field, positive charges moving in one direction, negative charges in the other. The molecules behave in some degree like conductors; only, since the positive and negative charges are bound together in the molecule, they cannot move far apart, and there is no flow of current. According to the second possibility, we think of the molecules as permanent dipoles (as in most cases they are) which develop a certain measure of orientation under the influence of the applied field acting in opposition to the disordering influence of thermal agitation. Such an orientation process must clearly be governed by the Boltzmann distribution law and would result in the dielectric constant's being strongly dependent on the absolute temperature; as the system becomes hotter, the random motion of the molecules increases, and the electric field becomes less effective in orienting them. In contrast, if the polarization were mainly the result of the first process, a displacement of charge, then the dielectric constant at constant volume should be essentially independent of temperature; for in this case the molecules are behaving like metallic conductors when subjected to an electric field. In general we may expect both possibilities to be realized, and experiment shows that this is so. For certain substances, always those of low dielectric constant whose molecules may on structural grounds be expected to possess no appreciable permanent moment, the dielectric constant is essentially independent of temperature and nearly equal to its optical value given, as we shall see shortly, by the square of the index

of refraction. For other substances, which are always those of higher dielectric constant and for whose molecules we predict an appreciable permanent moment, the dielectric constant decreases markedly with rise of temperature, at constant volume, and is always greater than the square of the index of refraction. Clearly, it should be possible, by implementing these ideas, to obtain much information about molecular structure from the study of the dielectric constant, but in order to do so it is necessary to give them a quantitative formulation. The first such formulation, which is basic to all further developments, is due to Debye.

In order to present Debye's theory, we begin by writing the expression for the electric moment, \mathbf{I} , per unit volume, developed by the medium, in terms of n , the number of molecules per unit volume, and $\bar{\mathbf{m}}$, the mean moment of a molecule *in the direction of the field*. This is simply

$$\mathbf{I} = n\bar{\mathbf{m}} \quad (10)$$

We now assume that $\bar{\mathbf{m}}$ is proportional to the electric field strength, \mathbf{F} , acting on a molecule, setting

$$\bar{\mathbf{m}} = \alpha\mathbf{F} \quad (11)$$

where α is a constant of proportionality. This is justifiable for values of \mathbf{F} which are not too large—not large enough to produce what are called saturation effects—as subsequent analysis will show. We use \mathbf{F} instead of \mathbf{E} to denote the internal field strength acting on a molecule, since in general we cannot expect this to be the same as the macroscopic electric intensity. The subsequent development of the theory consists of two parts, one of which is concerned with \mathbf{F} and the other with α .

It will be noted that if only \mathbf{F} were the *same* as \mathbf{E} , then, since

$$4\pi\mathbf{I} = (D - 1)\mathbf{E}$$

$n\alpha$ could be equated directly to the electrical susceptibility $(D - 1)/4\pi$. Actually we cannot expect anything so simple, for the polarized state of surrounding molecules will contribute to the force acting on any molecule in the interior of a dielectric. In his analysis of \mathbf{F} , Debye follows the course of earlier investigators, notably Lorentz, who were concerned mainly with optical problems. He assumes that the dielectric fills the space between the plates of a parallel plate condenser so charged that the electric intensity is \mathbf{E} , and introduces the conceptual device of dissecting out a small sphere of the dielectric—a sphere large in terms of molecular dimensions but small in terms of macroscopic standards—leaving behind only the molecule under consideration at the center of the spherical cavity. The electric intensity \mathbf{F} acting on this molecule is then considered to be the sum of three components: \mathbf{F}_1 , which results from the charges $+\sigma$ and

$-\sigma$ on the condenser plates and the charges $-\sigma'$ and $+\sigma'$ on the faces of the insulator immediately adjoining them; \mathbf{F}_2 , which is the force due to the induced charge on the walls of the spherical cavity; and \mathbf{F}_3 , which is the force arising from the polarized state of the molecules originally contained within the cavity. Component \mathbf{F}_1 is set equal to the electric

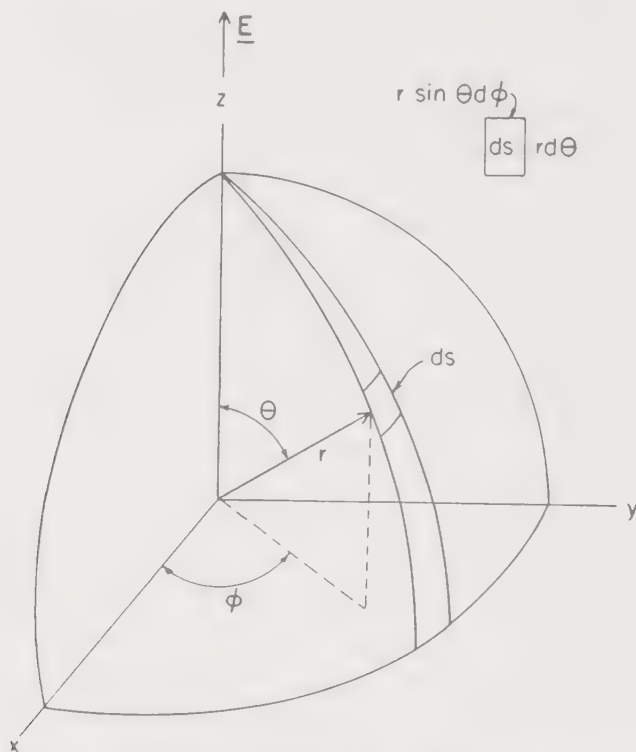


FIG. 3. The system of coordinates employed in describing the orientation of dipoles in an electric field. The small diagram at the upper right shows the alignment of surface area, dS , as the product of the length of its side.

intensity \mathbf{E} (see equations 7 and 8); \mathbf{F}_2 is calculated by a perfectly straightforward procedure to be given presently; and \mathbf{F}_3 is neglected. Strictly speaking, the neglect of \mathbf{F}_3 is justifiable only for a cubic crystal, but as an approximation it is allowable also for gases, nonpolar liquids, and dilute solutions in nonpolar liquids. By neglecting \mathbf{F}_3 , Debye limits his theory to such cases.³

For the calculation of \mathbf{F}_2 , we introduce a system of spherical coordinates with origin at the center of the cavity and with axis parallel to \mathbf{E} (Fig. 3). Then, by equation (9), the density of charge at any point on the surface of the cavity, which depends only on θ , is given from (9) by

³ The subdivision of the total electric intensity into three components, as adopted here, differs slightly from that in the presentation by Debye (1929). Our calculation is essentially identical with his, however, and the result is of course the same.

$\sigma' = I \cos \theta$. If r is the radius of the sphere, the force acting on a unit charge at the center of the sphere, due to the charge on an element of surface $ds = r^2 \sin \theta d\theta d\phi$, is therefore, using equation (9) and Coulomb's law, equal to

$$\frac{Ir^2 \cos \theta \sin \theta d\theta d\phi}{r^2}$$

and its component in the \mathbf{E} direction is this same quantity multiplied by $\cos \theta$. Since r^2 cancels out, it follows that the total field intensity \mathbf{F}_2 due to the charge on all elements of the surface (which for reasons of symmetry must be parallel to \mathbf{E}) is

$$\mathbf{F}_2 = \mathbf{I} \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} \cos^2 \theta \sin \theta d\theta d\phi = \frac{4}{3}\pi\mathbf{I} \quad (12)$$

The total value of \mathbf{F} (neglecting \mathbf{F}_3) is, therefore,

$$\mathbf{F} = \mathbf{F}_1 + \mathbf{F}_2 = \mathbf{E} + \frac{4\pi\mathbf{I}}{3} \quad (13)$$

By introducing this into (11) and making use of (10) and (7) to eliminate $\bar{\mathbf{m}}$, \mathbf{E} and \mathbf{I} , we arrive at the result

$$\frac{D-1}{D+2} = \frac{4}{3}\pi n\alpha \quad (14)$$

This is the famous Clausius-Mossotti relation. It was first introduced in other connections long before the advent of the Debye theory, but its retention there constitutes a principal feature of that theory.

In order to complete the theory, it is necessary to express α in terms of molecular properties. To accomplish this, we begin by writing it as the sum of two parts, α_0 and α_1 , one for each of the two kinds of polarization we have assumed. We identify α_0 with the first of these, that is the polarization resulting from a displacement of charge within the molecules, and give it no further analysis; α_1 then corresponds to the polarization due to the orientation of the molecules as permanent dipoles. Similarly, we break up the mean moment, $\bar{\mathbf{m}}$, of a molecule in the direction of the field into two parts, $\bar{\mathbf{m}}_0$ and $\bar{\mathbf{m}}_1$, one for each of the two kinds of polarization corresponding to α_0 and α_1 , respectively, setting

$$\begin{aligned} \bar{\mathbf{m}}_0 &= \alpha_0 \mathbf{F} \\ \bar{\mathbf{m}}_1 &= \alpha_1 \mathbf{F} \end{aligned}$$

If now we denote the permanent moment of a simple molecule by μ ,

then we also have the relation

$$\bar{m}_1 = \mu \overline{\cos \theta}$$

where $\overline{\cos \theta}$ is the mean value of the cosine of the angle θ between the dipole axis of a molecule and the direction of the field; $\overline{\cos \theta}$ represents the degree of orientation of the molecules under the influence of the field and may be calculated from the Boltzmann distribution law.

The orienting force acting on a molecule and tending to line it up in the direction of the field is $\mu F \sin \theta$. The energy of a molecule is obtained from the potential of this force (see p. 255) and is

$$W = -\mu F \cos \theta$$

It follows that the number of molecules having any given orientation, θ , with respect to the field is proportional to

$$e^{\mu F \cos \theta / kT}$$

Here θ may be identified with the polar angle of a system of spherical coordinates whose axis is parallel to the field. The orientation of any molecule is given by the values of the two angles, θ and ϕ , of this system. The number of molecules whose orientation falls within the interval between ϕ and $\phi + d\phi$ and between θ and $\theta + d\theta$ is, therefore, proportional—see Fig. 3, noting that $r = 1$ in this case—to

$$e^{\mu F \cos \theta / kT} \sin \theta d\theta d\phi$$

this being the product of the Boltzmann distribution function and an element of surface on a unit sphere corresponding to the interval. The mean value $\overline{\cos \theta}$ is obtained by multiplying the number of molecules having the orientation defined by the angle θ by the corresponding value of $\cos \theta$, integrating this expression over all values of θ and ϕ , and dividing by the total number of molecules:

$$\overline{\cos \theta} = \frac{\int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} e^{\mu F \cos \theta / kT} \cos \theta \sin \theta d\theta d\phi}{\int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} e^{\mu F \cos \theta / kT} \sin \theta d\theta d\phi} \quad (14.1)$$

The integral in the denominator of this expression, giving the total number of molecules in the system, is

$$\frac{1}{a} (e^a - e^{-a}) = \frac{1}{a} \sinh a$$

where $a = \mu F / kT$. The integral in the numerator may be obtained most easily by noticing that it is, from a mathematical point of view, the

derivative of that in the denominator with respect to $\mu F/kT = a$ as variable. This is

$$\frac{1}{a} (e^a + e^{-a}) - \frac{(e^a - e^{-a})}{a^2} = \frac{\cosh a}{a} - \frac{\sinh a}{a^2}$$

The quotient of the two, which gives $\overline{\cos \theta}$, is

$$\overline{\cos \theta} = \coth a - \frac{1}{a}$$

This expression is known as the Langevin function. It was first derived by the French physicist Langevin in his studies on magnetism. We shall

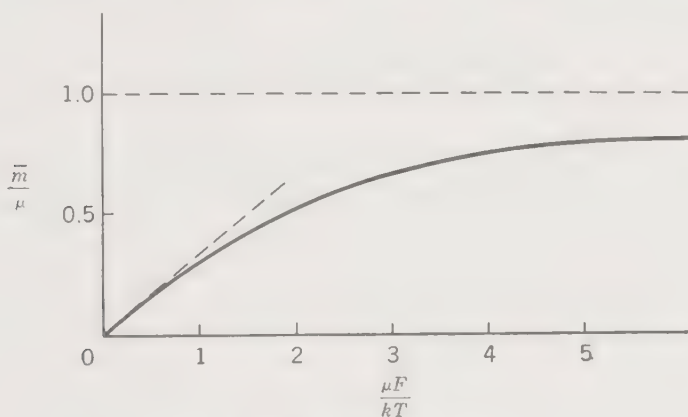


FIG. 4. The Langevin function.

denote it by $L(a)$. As a varies from 0 to infinity, $L(a)$ increases from 0 to 1 (Fig. 4). For small values of the argument it may be replaced by the approximation

$$L(a) \cong \frac{a}{3}$$

If we assume, therefore, that $\mu F/kT$ is small in relation to unity, we can write

$$\overline{\cos \theta} = \frac{\mu F}{3kT}$$

and consequently

$$\bar{m}_1 = \frac{\mu^2 F}{3kT}$$

This corresponds to the assumption introduced earlier, in the analysis of F , that for small values of F the moment of a molecule can be regarded as proportional to F . As F becomes large, and $L(a)$ approaches its upper limit 1, this assumption breaks down. We encounter a condition of "sat-

uration" in which nearly all the molecules are lined up in the field and a further increase of F produces relatively little effect. Physical situations in which this condition is realized are seldom encountered in the study of dielectric constants.

If we forget saturation effects and content ourselves with the approximation $L(a) = a/3$, it follows that

$$\alpha_1 = \frac{\bar{m}_1}{F} = \frac{\mu^2}{3kT} \quad (15)$$

This relation is a second fundamental feature of the Debye theory, the first being the Clausius-Mossotti relation. If we combine the two, we obtain at once as the full embodiment of the theory the equation

$$\frac{D-1}{D+2} = \frac{4}{3} \pi n \left(\alpha_0 + \frac{\mu^2}{3kT} \right) \quad (16)$$

The quantity on the right, which is often designated by p , is known as the polarizability per unit volume, or sometimes, though less appropriately, as the polarization per unit volume. It is dimensionless. The related quantity

$$\frac{4}{3} \pi \left(\alpha_0 + \frac{\mu^2}{3kT} \right)$$

applicable to a single molecule, is known as the molecular polarizability. Since n gives the number of molecules per unit volume, it will be seen from (16) that this has the dimensions of volume. For many purposes, it is convenient to introduce the molar polarizability, obtained by multiplying the molecular polarizability by Avogadro's number. This also has the dimensions of volume. It is generally denoted by P . It will be seen that

$$P = \frac{D-1}{D+2} \cdot \frac{N}{n} = \frac{D-1}{D+2} \cdot V = \frac{D-1}{D+2} \frac{M}{\rho} \quad (17)$$

where M is the molecular weight, V the molar volume, and ρ the density. In dealing with some problems, it is also helpful to make use of still another variant, namely the polarizability per unit mass. This is given by the volume polarizability divided by the density (ρ) and has the dimensions of a reciprocal density. It may be expressed as

$$p' = \frac{D-1}{D+2} \frac{1}{\rho} = \frac{4}{3} \pi \left(\alpha_0 + \frac{\mu^2}{3kT} \right) n' \quad (18)$$

where n' is the number of molecules per unit weight.

It will be seen that the volume polarizability of a medium is additive

in the molecular polarizabilities. Therefore, if we consider a system containing i species of molecules and denote by n_i the number of molecules of the i th species per unit volume,

$$\frac{D-1}{D+2} = \frac{4}{3} \pi \sum_i n_i \alpha_i \quad (19)$$

where of course

$$\alpha_i = \alpha_{0i} + \frac{\mu_i^2}{3kT}$$

Alternatively, the volume polarizability may be expressed in terms of the i molar polarizabilities, P_i , as

$$\frac{D-1}{D+2} = \sum_i N_i P_i \quad (20)$$

when N_i denotes the number of moles of i th species per unit volume. These simple relations are fundamental in dealing with mixtures and solutions.

Application of the Debye Theory

According to the Debye theory the molar polarizability of a substance, defined by

$$P = \frac{4}{3} \pi N \left(\alpha_0 + \frac{\mu^2}{3kT} \right) \quad (21)$$

is given in terms of the dielectric constant by

$$P = \frac{D-1}{D+2} V = \frac{D-1}{D+2} \frac{M}{\rho} \quad (21a)$$

where N is Avogadro's number, V the molar volume, M the molecular weight, and ρ the density. A critical test of the theory is, therefore, provided by a plot of $V(D-1)/(D+2)$ against $1/T$. The theory predicts that such a plot will be a straight line, of which the slope is

$$\left(\frac{dP}{d(1/T)} \right) = \frac{4\pi N \mu^2}{9k} \quad (22)$$

and the intercept with the ordinate axis

$$\frac{4\pi}{3} N \alpha_0 \equiv P_0 \quad (23)$$

It is a clear vindication of the theory that plots of this kind based on dielectric constant measurements of gases are in fact actually straight

lines and that the slope of such lines is always either positive or zero, the latter event being realized in the case of highly symmetrical molecules, as well as some others, for which we have reason to expect that the dipole moment is zero. Plots of this kind are shown in Fig. 5. The same behavior is encountered when we consider the molar polarizabilities of substances in dilute solution in nonpolar solvents like benzene and hexane.⁴ There are, to be sure, occasional exceptions where the straight line

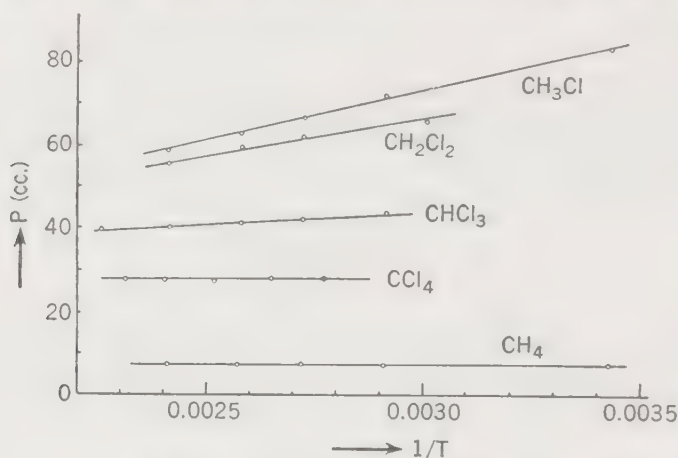


FIG. 5. Polarization as a function of $1/T$ for some simple compounds. Note that the symmetrical compounds CH_4 and CCl_4 give horizontal curves, indicating zero dipole moment. The other compounds, with finite dipole moments, give lines of positive slope. From Debye (1929), p. 39.

consists of two sections, of different slopes, separated by a break, but these rare cases are always susceptible of some reasonable explanation, such as a change of moment associated with the sudden onset of free rotation at a certain critical temperature. In view of these considerations and because of the consistency of the whole picture which it provides, there can be no doubt therefore of the fundamental correctness of the Debye theory as applied to gases and dilute solutions in nonpolar solvents.

It will be recalled that α_0 represents the polarizability of a molecule due to the displacement of charge. It may be supposed to arise from two

⁴ The calculation of the polarizability of a substance in solution involves breaking up total measured polarizability into two parts, one for the solvent, one for the solute. Usually the simplest way to do this is to make use of the polarizability per unit mass, which we denote by p' . If w_1 and $w_2 = (1 - w_1)$ represent the weight fractions of solvent and solute, respectively, then we write

$$\frac{D - 1}{D + 2} \frac{1}{\rho} = p' = w_1 p'_1 + w_2 p'_2$$

From two measurements at different values of w_2 , or, ideally, from measurements on a dilution series, it is possible to separate p'_1 and p'_2 . The molar polarization, P , of any substance is, of course, given by $p'M$, where M is the molecular weight.

contributions, an atomic one due to the displacement of atoms, and an electronic one due to the displacement of subatomic particles, of which the latter may be expected to be predominant. According to electromagnetic theory, the dielectric constant is equal to the square of the generalized index of refraction. For the very rapidly alternating field of ordinary light, we may expect only the electronic polarizability of the molecules to come into play—in any case there will be no time for the polar molecules to orient themselves in the field, even though there may be some question as to the displacement of the atoms. Clearly it is important to compare the optical polarizability calculated from equation (16), D being replaced by n^2 (the square of the optical index of refraction), with the polarizability obtained from the intercept of the curves of total polarizability versus $1/T$.⁵ Such a comparison shows that the agreement is uniformly good, though, as might be expected in view of the atom polarizability, it is found that the optical polarizability is nearly always slightly less than the other. The difference is so small, however, that for most purposes it may be neglected and $(\frac{4}{3})\pi N\alpha_0 \equiv P_0$ may be identified with the optical polarizability, $V(n^2 - 1)/(n^2 + 2)$. This provides additional confirmation of the theory and is also of practical significance for the calculation of electric moments.

Ideally the best way to determine the electric moment of a substance is from the slope of the curve of total polarizability plotted against $1/T$. This demands a series of dielectric constant measurements over as wide a temperature range as possible. In practice, in view of what has been said in the last paragraph, it is easier and almost as satisfactory to base the calculations on a single determination of the total polarizability (dielectric constant) obtained at some one temperature and the value of the optical polarizability obtained from a measurement of the index of refraction. This has become common practice.

Over the course of the last twenty-five years, since the appearance of the Debye theory, a vast amount of data on the dipole moments of a wide variety of substances has been accumulated. Certain representative results are given in Table I. It is to be noted that they are all stated in terms of what is known as the Debye unit, which requires a brief explanation. The molar polarizability, $P = V(D - 1)/(D + 2)$, has the dimensions of volume. From (21) and (23)

$$\mu = \sqrt{\frac{9kT}{4\pi N}} (P - P_0) \quad (24)$$

⁵ In making the comparison it is desirable to use the index of refraction for a wavelength far enough from the position of an absorption band to avoid the disturbing influence of the band.

TABLE I
 DIPOLE MOMENTS (μ) OF SOME IMPORTANT MOLECULES
 (All moments are given in Debye units: 1 Debye unit = 10^{-18} esu)

Molecule	Formula	μ	Physical state*
Oxygen	O ₂	0	Gas
Water	H ₂ O	1.84	G
Hydrogen peroxide	H ₂ O ₂	2.06	Dioxane
Hydrogen chloride	HCl	1.08	G
Ammonia	NH ₃	1.46	G
Carbon dioxide	CO ₂	0	G
Methane	CH ₄	0	G
Ethane	C ₂ H ₆	0	G
Methyl acetylene	C ₃ H ₄	0.72	G
Cyclohexane	C ₆ H ₁₂	0	Benzene
Benzene	C ₆ H ₆	0	G
2,2,3-Trimethylbutane	C ₇ H ₁₆	0	Fl
Toluene	C ₇ H ₈	0.37	G
Methyl chloride	CH ₃ Cl	1.86	G
Methyl bromide	CH ₃ Br	1.80	G
Methyl iodide	CH ₃ I	1.59	G
Chloroform	CHCl ₃	1.02	G
Bromoform	CHBr ₃	1.3	G
Iodoform	CHI ₃	1.00	Hexane
<i>o</i> -Dichlorobenzene	C ₆ H ₄ Cl ₂	2.59	G
<i>m</i> -Dichlorobenzene	C ₆ H ₄ Cl ₂	1.67	G
<i>p</i> -Dichlorobenzene	C ₆ H ₄ Cl ₂	0	G
Methyl alcohol	CH ₃ OH	1.705	G
Formaldehyde	CH ₂ O	2.17	G
Formic acid	CH ₂ O ₂	2.07	Dioxane
Dimethyl ether	C ₂ H ₆ O	1.29	G
Acetaldehyde	C ₂ H ₄ O	2.68	G
<i>n</i> -Propyl alcohol	C ₃ H ₈ O	1.66	G
Acetone	C ₃ H ₆ O	2.84	G
Diethyl ether	C ₄ H ₁₀ O	1.14	G
Ethyl acetate	C ₄ H ₈ O ₂	1.76	G
Phenol	C ₆ H ₆ O	1.40	G
Methylamine	CH ₅ N	1.33	G
Dimethylamine	C ₂ H ₇ N	1.02	G
Trimethylamine	C ₃ H ₉ N	0.62	G
Pyridine	C ₅ H ₅ N	2.21	Benzene
Aniline	C ₆ H ₇ N	1.48	G
Nitromethane	CH ₃ O ₂ N	3.50	G
Urea	CH ₄ ON ₂	4.56	Dioxane
Acetamide	C ₂ H ₅ ON	3.6	Benzene
<i>sym</i> -Dimethylurea	C ₃ H ₈ ON ₂	4.8	Dioxane
<i>p</i> -Nitroaniline	C ₆ H ₆ O ₂ N ₂	6.3	Benzene
<i>m</i> -Dinitrobenzene	C ₆ H ₄ O ₄ N ₂	4.07	Hexane
<i>p</i> -Dinitrobenzene	C ₆ H ₄ O ₄ N ₂	0	Benzene

* G denotes a dipole moment measured in the vapor state; the words Dioxane, Benzene, Hexane, denote solutions in these liquids.

This expression, which gives μ in electrostatic units (esu), has, therefore, the dimensions of $M^{1/2}t^{-1}L^{3/2}$ when M , t , and L stand for mass, time, and length, respectively. If we base our calculations on the second, the gram, and the centimeter, the P 's will be expressible in cubic centimeters and μ will be given numerically by

$$\mu = 0.0127 \times 10^{-18} \sqrt{(P - P_0)T} \text{ esu} \quad (25)$$

By omitting the factor 10^{-18} we express μ (without changing its fundamental dimensions) in terms of a new unit known as the Debye unit. One Debye unit is 10^{-18} esu.

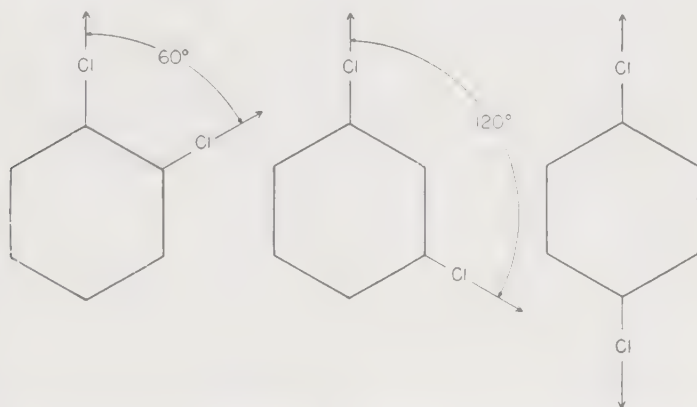


FIG. 6. The three dichlorobenzenes, illustrating the vector addition of the dipole moments of the two C—Cl bonds in each molecule.

One of the interesting features of these data is the behavior of the three dichlorobenzenes—*ortho*, *meta*, *para*. This provides a good illustration of the principle of the vector additivity of polar groups in a rigid molecule. Benzene itself is nonpolar, but the chloro group has a large moment (generally considered to have its negative end associated with the chlorine atom). In the case of *p*-dichlorobenzene the two polar groups would be expected to point in opposite directions and cancel one another. The zero moment of this molecule shows that this expectation is indeed realized. On the other hand, in the *ortho* compound the two groups point more or less in the same direction, the angle between them being ideally 60° in the absence of steric effects. This corresponds with the relatively large observed moment. The *meta* compound should lie somewhere in between, as it does (Fig. 6).

Benzene is a rigid planar molecule, owing to the resonating system of double bonds which it contains. In other molecules, more particularly very large molecules where there is something approaching free rotation of component parts with respect to one another, the situation is quite different, and the principle of the vector additivity of moments does not

hold. Hence we may expect the constituent dipoles to orient more or less independently of one another, as if they were separate molecules. To the extent to which this is true it will be the polarizabilities and not the moments which are additive. (Compare equation 20 for separate molecules.) Since polarizability is proportional to the square of the moment, this means that the square of the total mean moment of such a molecule will be the sum of the squares of the component moments. (When we are concerned with a nonrigid molecule in which a number of different configurations are realized, the quantity which we measure is a mean polarizability, or a mean moment.) In contrast, in rigid molecules like the dinitrobenzenes, the square of the moment is equal to the square of the vector sum of the component moments.

TABLE II
POLARIZABILITIES PER GRAM OF HYDROXYDECANOIC ACID POLYMERS

Mean molecular weight	Mean number of monomer units	p'^*
1715	10	0.411
3190	19	0.410
4170	24	0.417
5670	33	0.383
9330	55	0.415
28650	168	0.415

* Averaged from values at several concentrations.

A good example of a molecule in which the component polar groups orient essentially independently of one another is provided by the hydroxydecanoic acid polymers synthesized by W. H. Carothers. Hydroxydecanoic acid consists of a straight chain of ten carbon atoms with a hydroxyl group at one end and a carboxyl group at the other. There are no double bonds in the chain. Polymerization results from the formation of ester linkages between hydroxyl and carboxyl groups. These ester linkages are the only polar groups (except for the terminal groups) in a polymer. The data given in Table II were obtained on polymer fractions of different mean molecular weights, as indicated, in dilute solution in the nonpolar solvent benzene. It will be seen that the polarizability per unit mass is constant over the whole range of molecular weight studied, or in other words that the polarizability is proportional to the molecular weight. This implies essentially independent orientation of the component polar groups. The hydroxydecanoic acids may be thought of as the prototype of one kind of large molecule, characterized by flexibility. On the other hand, many proteins, as we shall see later, stand in marked contrast and seem to represent the other extreme of rigid molecules.

It would be easy to continue the discussion of the important information and ideas which emerge from the application of the Debye theory to dielectric constant data on gases and dilute solutions in nonpolar solvents. Space and the limitations of our subject prevent our doing so; the reader may be referred to the monograph by Debye himself (1929) and to a more recent one by Smyth (1955).

The Breakdown of the Debye Theory in Polar Liquids

Although there can be no doubt that the Debye theory is essentially correct when applied to gases and dilute solutions in nonpolar solvents, there can likewise be no doubt that it rapidly loses validity as we depart from such systems and that it breaks down completely in the case of strongly polar media like liquid water. One of the most compelling arguments in this connection rests on the very form of the basic equation of the theory. According to this the molar polarizability can never exceed the molar volume, for this is the limiting value approached by $V(D - 1)/(D + 2)$ as the dielectric constant becomes infinite. On the face of it this arbitrary limitation is unreasonable, if not indeed absurd, for the molar volume is surely unconnected with the electric moment of the molecules. It looks even more unreasonable when we consider specific cases. Thus, for example, the molar polarization of water calculated on the basis of (17) from the dielectric constant of the liquid (78.54 at 25°) is about 17 cc. Even if the dielectric constant were infinite, it would be only slightly greater—about 18 cc. In contrast, the value calculated from the dielectric constant of the vapor, where we know the theory to be reliable, is close to 100—i.e., about five times as great.

Another closely related objection to the theory rests on the concept of the Curie point. This concept was introduced by Pierre Curie long ago in connection with studies of magnetism, but it is directly applicable, like the Langevin function, to the study of dielectric constants. If we solve (16) for D in terms of the volume polarizability, p , we obtain

$$D = \frac{2p + 1}{1 - p} \quad (26)$$

The value of p itself, however, is given by

$$p = \frac{4\pi n}{3} \left(\alpha_0 + \frac{\mu^2}{3kT} \right)$$

It is evident that, as T diminishes, p will increase, partly because of an increase of n associated with increasing density, but, more significantly, because of the presence of $1/T$ as multiplier of μ^2 . At a certain critical

temperature p will become equal to unity, and then D should go to infinity. This temperature, if it exists, is known as the Curie point. Something like a Curie point does, in fact, turn up in certain rare cases involving the dielectric constant. Mostly, however, it fails to materialize as it ought if (26) were valid. It may be remarked that the sudden increase of D toward infinity predicted by the theory at a certain critical value of $\mu^2/3kT$ has been rather dramatically referred to by Van Vleck as the “ $4\pi\mathbf{I}/3$ catastrophe,” since, as analysis of the derivation of (16) reveals, it arises from the presence of the term $4\pi\mathbf{I}/3$ in the expression for F . This is significant, since it points to a possible source of our difficulties.

Still another indication of the breakdown of the Debye theory, and one which suggests the form which any alternative theory applicable to polar media should take, results from an empirical consideration of the dielectric properties of a large number of polar liquids. A correlation of the moments of such molecules (as determined from measurements on the vapor or in dilute solutions in nonpolar solvents) with their dielectric constant in the liquid state shows that the relation between dielectric constant and polarizability (calculated from the moments by 21) is essentially linear instead of that required by the Debye theory. According to the latter, a plot of D versus p should give a hyperbola with asymptotes at $p = 1$ and $D = -2$ (see equation 26). The same fundamental linearity between dielectric constant and polarizability (or at least something proportional to the square of the dipole moment) is also revealed by a large body of data on the dielectric properties of solutions of dipolar ions. All these data are discussed in detail later in this chapter.

At this point we pass on to a consideration of modifications of the Debye theory, or alternatives to it, which may be applicable to strongly polar media.

Modifications of the Debye Theory

As we have just seen, one of the telling objections to the Debye theory involves the so-called “ $4\pi\mathbf{I}/3$ catastrophe,” which can be traced back to the occurrence of the term $4\pi\mathbf{I}/3$ in the expression for F , the actual force acting on a molecule in the interior of a dielectric. It behooves us, therefore, in looking for a reasonable modification of the theory, to reconsider the analysis of F .

It will be recalled that the equation

$$\mathbf{F} = \mathbf{E} + \frac{4}{3}\pi\mathbf{I}$$

which forms such a fundamental part of the Debye theory, and leads to the Clausius-Mossotti relation, results from the conceptual procedure of dissecting out a small sphere of dielectric, leaving behind only the mole-

cule under consideration at its center, and expressing \mathbf{F} in terms of the free charges on the walls of the cavity so formed and on the plates of a condenser between which the dielectric is supposed to be contained. The contribution to \mathbf{F} arising from the material contained in the small sphere is neglected. A possible objection to this procedure is that in calculating the charges on the wall of the cavity we proceed as if the electric intensity were unaffected by the presence of the cavity, although in fact a cavity distorts the lines of force in its vicinity. In answer to this, it might be maintained that the cavity is, after all, only a fiction, and that the small sphere of dielectric should not be thought of as being actually removed but only as being neglected in its influence on the central molecule. The question thus raised has no clear answer, for as we approach molecular dimensions the concept of the dielectric constant loses definition. It suggests, however, that at least we explore the result of treating the cavity as a real cavity which acts to modify the field. This we now proceed to do. The modified field includes the influence of the polarization charge on the walls of the cavity.

Here we anticipate in some degree a model which we shall meet later in this chapter in a much more elaborate treatment of dielectrics due to Onsager. We picture the polar molecule as a dipole, embedded in a spherical cavity of low dielectric constant. We assume also that the surrounding medium is continuous and characterized by the macroscopic value of the dielectric constant right up to the surface of the sphere. We do not specify the exact size of the cavity except to require that it be small enough to exclude other dipoles, which might exert a significant effect on the molecule under consideration, nor do we specify the nature of the dipole or its position within the sphere. The effect of the sphere of low dielectric constant will be to disturb the distribution of the lines of force with a resulting modification in the electric field everywhere in its vicinity, both inside and outside the sphere. We are, of course, primarily interested in the field in the interior of the sphere, for it is this which acts on the dipole.

The problem is to calculate the effect of introducing a sphere of radius a and dielectric constant D_1 into an indefinitely extended region of dielectric constant D_2 , on the electric intensity \mathbf{E} which was originally uniform and of value \mathbf{E}_1 throughout the region. This is a problem of classical electrostatics. It may be solved in terms of the potential, ψ , of \mathbf{E} . Since there are no local charges anywhere in the region, ψ must satisfy Laplace's equation ($\nabla^2\psi = 0$). It must also conform to the boundary conditions, which are that ψ be everywhere finite and continuous, in particular at the surface of the sphere, and that the normal component of the electric displacement (\mathbf{D}) be the same on both sides of that surface, where the dielectric constant suddenly changes from D_1 to D_2 . (This last condition rep-

resents the fact that the surface of the sphere is uncharged. See equation 6. In this treatment we have only to consider actual charges, not polarization charges.)

The potential, ψ , may be expressed as the sum of two parts, ψ_1 , the potential of the unperturbed electric intensity \mathbf{E}_1 which existed before the introduction of the sphere, and ψ_2 , the perturbation potential due to the presence of the sphere:

$$\psi = \psi_1 + \psi_2 \quad (27)$$

To describe the system (see Fig. 7) we introduce a system of spherical coordinates with origin at the center of the sphere and polar axis (which we identify with the x axis in Cartesian coordinates) antiparallel to \mathbf{E}_1 .

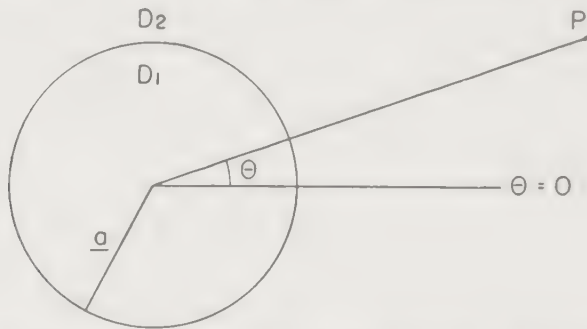


FIG. 7. A sphere of dielectric constant D_1 , immersed in a medium of dielectric constant D_2 , in an electric field.

Then, since the equipotential surfaces of ψ_1 are planes perpendicular to \mathbf{E}_1 , ψ_1 is independent of the azimuthal angle, ϕ , and at any point is given by $\psi_1 = E_1 x$ in Cartesian coordinates, or in polar coordinates (r, θ) by

$$\psi_1 = E_1 r \cos \theta \quad (28)$$

This, of course, satisfies Laplace's equation and is continuous throughout the whole region. The same must, therefore, be true of the perturbation potential, ψ_2 , which is also independent of ϕ for reasons of symmetry. The problem of determining ψ_2 is that of finding two solutions of Laplace's equation, one for the region outside the sphere, one for the region inside it, which conform to the boundary conditions. Now it is known from the theory of differential equations that continuous solutions of Laplace's equation in spherical coordinates are of the form

$$\psi = (Ar^l + Br^{-l-1})P_l^m(\cos \theta)e^{\pm im\phi} \quad (29)$$

where A and B are constants, and $P_l^m(\cos \theta)$ is what is known as an associated Legendre function, or an associated spherical harmonic, of the argument $\cos \theta$; l and m are integers, which, together with A and B ,

are to be chosen in accordance with the boundary conditions. In the present case, since ψ is independent of ϕ , m is equal to 0. This greatly simplifies the problem, for it means that the associated Legendre function degenerates into an ordinary Legendre function, P_l^0 . The forms of P_l^0 for various values of l are well known, e.g.:

$$P_0^0(\cos \theta) = 1; \quad P_1^0(\cos \theta) = \cos \theta; \quad P_2^0 = \frac{(3 \cos^2 \theta - 1)}{2}$$

It is unnecessary to consider larger values of l , for a choice from the functions just given suffices for our problem. Indeed, a solution applicable inside the sphere is given by

$$\psi_2 = Ar \cos \theta \quad (\text{inside}) \quad (30)$$

corresponding to $l = 1$ and $B = 0$; and a solution applicable outside it is given by

$$\psi_2 = \frac{B \cos \theta}{r^2} \quad (\text{outside}) \quad (31)$$

corresponding to $l = 1$ and $A = 0$. These equations provide for the requirement that the potential ψ_2 must remain finite everywhere within the sphere, and that outside the sphere it must vanish at infinity. The condition for the continuity of ψ_2 at the surface of the sphere ($r = a$) is that ψ_2 (inside) = ψ_2 (outside), and therefore

$$Aa^3 = B \quad (32)$$

The condition for the equality of the normal component of the total electric displacement on the two sides of the surface of the sphere is

$$D_1 \frac{\partial(\psi_1 + \psi_2)}{\partial r} (\text{inside}) = D_2 \frac{\partial(\psi_1 + \psi_2)}{\partial r} (\text{outside}) \quad (\text{at } r = a) \quad (33)$$

By making use of these two equations we obtain

$$A = \frac{(D_2 - D_1)E_1}{(2D_2 + D_1)} \quad (34)$$

$$B = a^3 \frac{(D_2 - D_1)E_1}{(2D_2 + D_1)} \quad (35)$$

The *total* potential at any point *inside* the sphere is therefore

$$\begin{aligned} \psi_{\text{inside}} = \psi_1 + \psi_2 = E_1 r \cos \theta + \left(\frac{D_2 - D_1}{2D_2 + D_1} \right) E_1 r \cos \theta \\ + \left(\frac{3D_2}{2D_2 + D_1} \right) E_1 r \cos \theta \end{aligned} \quad (36)$$

This is the potential of the total force, \mathbf{E} , effective inside the sphere due to the applied field, which at a point far removed from the sphere is \mathbf{E}_1 . It follows from (36) that \mathbf{E} , like \mathbf{E}_1 , is antiparallel to the polar axis and is given by

$$\mathbf{E} = \frac{3D_2}{(2D_2 + D_1)} \mathbf{E}_1 \quad (\text{inside}) \quad (37)$$

(This results from differentiating ψ with respect to r and setting $\cos \theta = 1$.) It should be noted that this result is independent of r , showing that \mathbf{E} is everywhere constant within the sphere. On the other hand, at any point external to the sphere the total potential is given by

$$\psi = E_1 r \cos \theta + \left(\frac{D_2 - D_1}{2D_2 + D_1} \right) a^3 E_1 \frac{\cos \theta}{r^2} \quad (\text{outside}) \quad (38)$$

which falls off to $E_1 r \cos \theta$ as $r \rightarrow \infty$. We have discussed the interpretation of this equation in terms of image charges in Chapter 5, p. 262.

Let us now apply the result given in equation (37) to the model which we set out to explore, according to which the molecule is treated as a dipole embedded in a sphere consisting of a medium of dielectric constant low in relation to that of the surroundings. The actual force, \mathbf{F} of the Debye theory acting on the dipole will then be identifiable with the \mathbf{E} of equation (37) and is, significantly enough, independent of the position of the dipole in the sphere. Also \mathbf{E}_1 and D_2 will be identifiable with the macroscopic electric intensity, \mathbf{E} , and the macroscopic dielectric constant, D . Consequently we rewrite (37) as

$$\mathbf{F} = \frac{3D}{2D + D_1} \mathbf{E} \quad (39)$$

In (39) D_1 continues to represent the dielectric constant of the sphere, as throughout.

When this expression for \mathbf{F} is combined with the two basic equations $\mathbf{I} = n\bar{\mathbf{m}} = n\alpha\mathbf{F}$ and $(D - 1)\mathbf{E} = 4\pi\mathbf{I}$, we obtain at once, as a substitute for the Clausius-Mossotti relation, the result

$$\frac{(2D + D_1)(D - 1)}{9D} = \frac{4}{3} \pi n\alpha = p \quad (40)$$

the term p being, as usual, the volume polarizability.⁶ Other parts of

⁶ Our primary interest in this model is its implications for the case where D , the macroscopic dielectric constant, is very large. If we desire the model to yield the Debye expression in the other extreme of a medium of very low dielectric constant,

the Debye theory involving the analysis of α may be supposed to remain unchanged.

The important thing about the result embodied in equation (40) is that the $4\pi\mathbf{I}/3$ catastrophe is avoided. As D increases indefinitely, the equation passes over into the limiting form

$$\frac{2D}{9} = \frac{4}{3} \pi n \alpha = p \quad (41)$$

and p and D become directly proportional to one another (compare with equation 26). Under such conditions the optical polarizability becomes a negligible fraction of the total, so that we may neglect α_0 in comparison with $\mu^2/3kT$. Consequently, equation (41) may also be written as

$$\frac{2D}{9} = \frac{4}{9} \pi n \frac{\mu^2}{kT} \quad (42)$$

This very simple result, based on the simple expedient of taking account of the modification of the applied field in the interior of a molecule by assuming it to be a sphere of low dielectric constant, accords well with the facts. Indeed it leads to values for the electric moments of highly polar molecules, such as dipolar ions, in highly polar solvents, which are at least of the right order of magnitude. Before we consider the experimental data, however, we shall discuss two much more refined and searching analyses of the subject, both of which lead to closely similar though more complicated results. One of these is due to Onsager, the other to Kirkwood.

The Onsager Theory

The Onsager theory of dielectrics involves a bold and drastic revision of the whole approach to the subject. Onsager starts by replacing the molecule under consideration by a spherical cavity of the dimensions of the molecule, at the center of which there is a point dipole having a moment equal to the total moment (permanent plus induced) of the

like a gas at low pressure, it is necessary to set D_1 , the internal dielectric constant of the cavity, equal to 1. Then as $D \rightarrow 1$

$$p \rightarrow \frac{D-1}{D+2}$$

This follows from multiplying the numerator and denominator of (40) by $D+2$ and taking the limit. This point is of interest in connection with Kirkwood's theory to be discussed later.

molecule (Fig. 8). Outside the cavity he assumes the medium to be continuous and characterized by a dielectric constant equal to the observed macroscopic dielectric constant, D , right up to the surface of the cavity. This surface is uncharged, except by polarization. Inside the cavity the dielectric constant is taken as unity, since it is in fact not much greater than unity, and the assumption that $D = 1$ simplifies the calculations. Nevertheless, the cavity is supposed to be polarizable, to a degree characterized by an internal index of refraction, n , so that under the influence

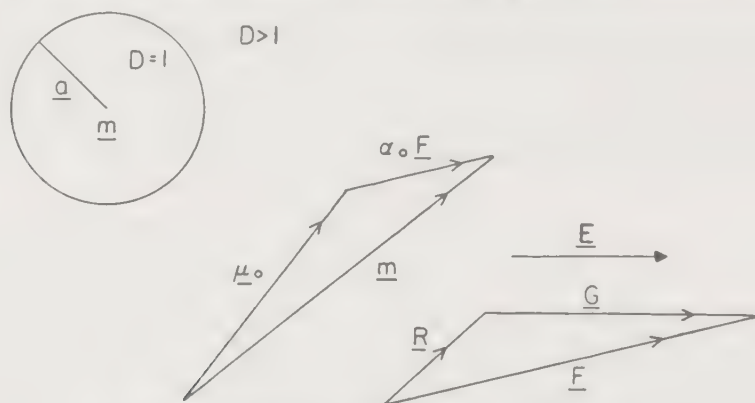


FIG. 8. Onsager's model of a spherical dipole.

of a force \mathbf{F} , it develops an induced moment, $\alpha_0 \mathbf{F}$, which is, of course, a vector parallel to \mathbf{F} . The term α_0 has the same meaning as the α_0 of the Debye theory and corresponds to a molecular polarizability, $(\frac{4}{3})\pi\alpha_0$, due to the displacement of charge. If we denote the radius of the cavity by a , it follows from the Clausius-Mossotti relation, which is retained for the calculation of this part of the polarizability, that

$$\alpha_0 = \left(\frac{n^2 - 1}{n^2 + 2} \right) a^3 \quad (43)$$

(for $(n^2 - 1)/(n^2 + 2)$ is the polarizability per unit volume given by that relation, and $4\pi a^3/3$ is the volume of the cavity). The orientation of the molecule is specified by a unit vector, \mathbf{u} , having the same direction and sense as its permanent moment \mathbf{u}_0 . It should be realized that this direction will not be the same as that of the total moment, which we denote by \mathbf{m} , since the latter comprises also the induced moment, $\alpha_0 \mathbf{F}$. The total moment, \mathbf{m} , is given by the vector sum⁷

$$\mathbf{m} = \mu_0 \mathbf{u} + \alpha_0 \mathbf{F}$$

⁷ It is of interest to compare this model with the less specific one introduced into the preceding section, in which the molecule was treated as a sphere of low dielectric constant with a dipole (not necessarily a point dipole) situated somewhere in its interior (not necessarily at its center). The much simpler calculations involved there made it possible to employ this more general model.

In order to proceed it is necessary to calculate \mathbf{F} . Everything hinges on this. Now it is assumed that \mathbf{F} is decomposable into two parts, one due directly to the applied field, \mathbf{E} , as modified by the presence of the cavity, and the other due to the polarization produced in the surrounding medium by the dipole itself. The latter is called the reaction field, since it represents the effect of the dipole reacting back on itself via the medium. It is clear that the reaction field will be peculiar to each molecule at any moment and will depend on the instantaneous orientation of the molecule. These two components of \mathbf{F} are denoted by \mathbf{G} and \mathbf{R} , respectively.

The evaluation of \mathbf{G} , the force due directly to the applied field, \mathbf{E} , presents no new problem. In fact the expression for \mathbf{G} may be written down at once on the basis of equation (37), which gives the force inside an insulating sphere of dielectric constant D_1 embedded in an indefinitely extended medium of dielectric constant D_2 , where the electric intensity at a point far removed from the sphere is \mathbf{E}_1 . All that is required is that we identify \mathbf{G} with \mathbf{E} , as given by (37), and \mathbf{E} —as employed in the Onsager theory—with \mathbf{E}_1 in (37). We also identify D with D_2 , and set $D_1 = 1$. The result is

$$\mathbf{G} = \frac{3D}{2D + 1} \mathbf{E} \quad (44)$$

Thus \mathbf{G} and \mathbf{E} are parallel vectors.

The reaction field, \mathbf{R} , can best be obtained from the expression for ψ , the potential anywhere due to the point dipole situated at the center of the cavity in the presence of its surroundings. The calculation of ψ follows essentially the same lines as that of the potential of an applied field as modified by the presence of an insulating sphere, which was given in the preceding section. We denote the moment of the point dipole by the vector \mathbf{m} and introduce a system of spherical coordinates with its origin in the dipole and the polar axis parallel to \mathbf{m} . As the expression for ψ both inside the cavity and outside it, we introduce the equation

$$\psi = \frac{m \cos \theta}{r^2} + \psi_1 \quad (45)$$

realizing, however, that ψ_1 will be given by different functions in each of the two regions. In this equation, the term $m \cos \theta / r^2$ gives the potential of the dipole in free space (see Chapter 5, equation 23) and takes care of the point singularity at $r = 0$, the site of the dipole, where the potential becomes infinite. Consequently, it follows that ψ_1 , whether inside or outside the cavity, is finite and continuous and is, therefore, a solution of Laplace's equation of the form given by equation (29) just as in the earlier

problem. Solutions applicable in the two regions are in fact the same as those given before, namely

$$\psi_1 = Ar \cos \theta \quad (\text{inside the cavity}) \quad (46)$$

$$\psi_1 = \frac{B \cos \theta}{r^2} \quad (\text{outside the cavity}) \quad (47)$$

The two constants A and B can be determined from the boundary conditions, also the same as those of the earlier problem, that the total potential, ψ , and the normal component of the electric displacement both be continuous at the surface of the cavity where the dielectric constant jumps from 1 to D . The result is

$$A = -\frac{2m(D-1)}{a^3(2D+1)} \quad (48)$$

$$B = -2m \frac{(D-1)}{(2D+1)} \quad (49)$$

When these values of A and B are introduced into equations (46) and (47) we obtain as the expressions for ψ in (45):

$$\psi = \frac{m \cos \theta}{r^2} - \frac{2m}{a^3} \left(\frac{D-1}{2D+1} \right) r \cos \theta \quad (\text{inside}) \quad (50)$$

$$\begin{aligned} \psi &= \frac{m \cos \theta}{r^2} - 2m \frac{(D-1)}{(2D+1)} \frac{\cos \theta}{r^2} \\ &= \frac{3D}{(2D+1)} \frac{m \cos \theta}{Dr^2} \quad (\text{outside}) \quad (51) \end{aligned}$$

It is worth noting that according to equation (51) the dipole gives rise to the same potential externally in the surrounding medium as a point dipole in the same medium without the cavity having a moment,

$$\mathbf{m}^* = \frac{3D}{2D+1} \mathbf{m} \quad (52)$$

For this reason \mathbf{m}^* is called the external moment of the dipole. The factor $[3D/(2D+1)]$ measures the effect of the cavity on the potential due to the dipole and is always greater than unity. (Compare section on image charges in Chapter 5.) But we are primarily interested in the situation in the interior of the cavity as described by (50). There the force consists of two parts—one, corresponding to $m \cos \theta/r^2$, which arises directly from the dipole, and one, corresponding to the second term (ψ_1), arising indirectly from it via the medium. The first cannot, of course, act on the molecule, but the latter can, and is in fact the force in which we are interested and which we have called the reaction field.

R. It will be seen that **R** acts parallel to the point dipole and, according to our convention regarding the coordinate system, in the same sense, i.e., so as to increase **m** by polarization. It is given by

$$R = - \left(\frac{\partial \psi_1}{\partial r} \right)_{\theta=0} = \frac{2m}{a^3} \left(\frac{D-1}{2D+1} \right) \quad (53)$$

and is constant throughout the cavity.

With the completion of the analysis of **F**, we are now in a position to express the moment of the molecule in terms of other quantities. To do this, we make use of the two equations:

$$\mathbf{F} = \mathbf{G} + \mathbf{R} = \frac{3D}{2D+1} \mathbf{E} + \frac{2(D-1)}{a^3(2D+1)} \mathbf{m} \quad (54)$$

and

$$\mathbf{m} = \mu_0 \mathbf{u} + \alpha_0 \mathbf{F}$$

If between these we eliminate **F**, the result is

$$\mathbf{m} \left[1 - \frac{2(D-1)\alpha_0}{(2D+1)a^3} \right] = \mu_0 \mathbf{u} + \frac{3D}{2D+1} \alpha_0 \mathbf{E} \quad (55)$$

When α_0 is expressed in terms of a^3 and the internal index of refraction, n , in accordance with equation (43), this becomes

$$\mathbf{m} = \frac{(n^2+2)(2D+1)}{3(2D+n^2)} \mu_0 \mathbf{u} + \frac{D(n^2-1)}{2D+n^2} a^3 \mathbf{E} \quad (56)$$

If we introduce the definition

$$\mu = \frac{(n^2+2)(2D+1)}{3(2D+n^2)} \mu_0 \quad (57)$$

this may be rewritten as

$$\mathbf{m} = \mu \mathbf{u} + \frac{D(n^2-1)}{2D+n^2} a^3 \mathbf{E} \quad (58)$$

The problem which now presents itself is to make use of this result in connection with the Boltzmann distribution law to calculate the mean total moment, \bar{m} , of a molecule in the direction of the field **E**. Consideration of equation (58) shows that this is

$$\bar{m} = \mu \overline{(\cos \theta)} + \frac{D(n^2-1)}{2D+n^2} a^3 E \quad (59)$$

where θ is the angle between the permanent dipole moment, μ_0 , of a molecule and the applied field, and $\overline{\cos \theta}$ is the value of $\cos \theta$ averaged over all the molecules. We note that \mathbf{u} is a unit vector parallel to the permanent dipole axis, and $\overline{\cos \theta}$ is the mean component of \mathbf{u} in the direction of the field, averaged over all the molecules in the system.

Our task, therefore, is simply to evaluate the mean value of $\cos \theta$. In order to apply the Boltzmann distribution law, it is necessary to formulate the expression for the energy of a molecule in any orientation, \mathbf{u} , defined by θ . To do this, we begin by considering the torque, \mathbf{M} , acting on a molecule as a result of the total field, \mathbf{F} , and the total moment, \mathbf{m} . We know that $\mathbf{F} = \mathbf{G} + \mathbf{R}$ and that \mathbf{G} is parallel to \mathbf{E} and \mathbf{R} to \mathbf{m} . The portion of the torque due to \mathbf{R} must, therefore, be zero. The remaining part, due to \mathbf{G} , is given by

$$\mathbf{M} = \mathbf{m} \times \mathbf{G} \quad (60)$$

where $\mathbf{m} \times \mathbf{G}$ denotes the vector product (see Chapter 5, p. 255). If we introduce the value of \mathbf{m} given by (58) and the value of \mathbf{G} from (44), this product is found to be simply

$$\mathbf{M} = \mathbf{u} \times \mathbf{G} \mu$$

since the vector product $\mathbf{E} \times \mathbf{E}$ vanishes. In terms of \mathbf{E} , this is

$$\mathbf{M} = \mathbf{u} \times \mathbf{E} \left(\frac{3D}{2D+1} \right) \mu \quad (61)$$

or, if we revert to scalar quantities, the absolute magnitude of \mathbf{M} is

$$M = \left(\frac{3D}{2D+1} \right) \mu E \sin \theta \quad (62)$$

If we now introduce still another definition, namely

$$\mu^* = \frac{3D}{2D+1} \mu = \frac{D(n^2+2)}{2D+n^2} \mu_0 \quad (63)$$

this becomes

$$M = \mu^* E \sin \theta \quad (64)$$

The work of orientation, i.e., the orientational energy of the molecule in the field, is the potential of this orienting torque, or

$$W = -\mu^* E \cos \theta$$

The minus sign accords with the fact that the potential increases with θ ; i.e., the torque acts to reduce θ and align the dipoles in the direction

of E . On introducing this expression into the Boltzmann distribution function, we obtain

$$\overline{\cos \theta} = \frac{\int_0^\pi \cos \theta e^{\mu^* E \cos \theta / kT} 2\pi \sin \theta d\theta}{\int_0^\pi e^{\mu^* E \cos \theta / kT} 2\pi \sin \theta d\theta} \quad (65)$$

This may be evaluated in exactly the same way as equation (14.1), with the result that

$$\overline{\cos \theta} = L\left(\frac{\mu^* E}{kT}\right) \quad (66)$$

where L denotes the Langevin function. For small values of the argument, this reduces to

$$\overline{\cos \theta} = \frac{\mu^* E}{3kT}$$

The mean moment of a molecule in the direction of the field is, therefore, from (59)

$$\bar{m} = \left[\frac{\mu\mu^*}{3kT} + \frac{D(n^2 - 1)a^3}{2D + n^2} \right] E \quad (67)$$

In order to derive from this an expression involving the dielectric constant we introduce the electric moment per unit volume, \mathbf{I} . On the one hand, this is given by

$$4\pi\mathbf{I} = (D - 1)\mathbf{E}$$

and on the other hand by

$$\mathbf{I} = n\bar{m}$$

where n is the number of molecules per unit volume. We also introduce one further relation in order to eliminate a^3 . Since the volume of each molecule is $4\pi a^3/3$, it follows that, if we assume that the molecules completely fill up the space,

$$\frac{n4\pi a^3}{3} = 1 \quad (68)$$

By making use of these three equations, we obtain

$$D - 1 = \frac{4\pi n\mu\mu^*}{3kT} + \frac{3D(n^2 - 1)}{2D + n^2} \quad (69)$$

This result is subject to the inconvenience that both μ and μ^* are themselves functions of D , but this may be readily removed by introducing the

definitions of these two quantities given in (57) and (63). After doing so, we arrive at the final equation

$$\frac{(D - n^2)(2D + n^2)}{D(n^2 + 2)^2} = \frac{4\pi n\mu_0^2}{9kT} \quad (70)$$

This relates the dielectric constant and the index of refraction to the permanent moment of the molecules. It is the analog, in the Onsager theory, of equation (16) or equation (21) of the classical Debye theory.

When D is large in relation to n^2 , as in strongly polar liquids, equation (70) approaches as a limiting form

$$D = \frac{(n^2 + 2)^2}{2} \cdot \frac{4\pi n\mu_0^2}{9kT} \quad (71)$$

Thus, D tends to become linear in the square of the dipole moment, that is to say, in the polarization per unit volume calculated according to classical concepts from values of the dipole moment obtained under ideal conditions. Indeed, as Onsager points out, the slope of the curve calculated from equation (71) by taking reasonable values for the internal index of refraction ($1.275 < n < 1.64$) accords closely with the slope of the actual curve obtained by plotting the measured dielectric constant against values of polarizability per unit volume reckoned from known values of the moments for a large number of polar liquids (see section on the breakdown of the Debye theory). On the other hand, when $D \rightarrow 1$, corresponding to the case of gas at low pressure, the Onsager treatment yields the classical Debye equation, as will be seen from (69) and the definitions of μ and μ^* . Also, if we set $\mu = 0$ and consequently $D = n^2$, we obtain the Clausius-Mossotti relation.⁸

It is possible, though slightly complicated, to adapt the foregoing results to the case of a medium consisting of several different kinds of molecules, which we distinguish by subscripts 1, 2, . . . , i . Equation (67) may be applied to each of these separately, with the result

$$\bar{m}_i = \left[\frac{\mu_i \mu_i^*}{3kT} + \frac{D(n_i^2 - 1)a_i^3}{2(D + n_i^2)} \right] E \quad (71.1)$$

⁸ This can best be shown by throwing (69) into the form

$$\frac{D - 1}{D + 2} = \frac{4}{3} \pi n \left[\frac{\mu \mu^*}{(D + 2)kT} + \frac{3D(n^2 + 2)\alpha_0}{(2D + n^2)(D + 2)} \right]$$

where we make use of the relation:

$$\alpha_0 = \frac{3}{4\pi n} \left(\frac{n^2 - 1}{n^2 + 2} \right)$$

Here, corresponding to the definitions (57) and (63),

$$\mu_i = \frac{(n_i^2 + 2)(2D + 1)}{3(2D + n_i^2)} \mu_{i_0} \quad (72)$$

$$\mu_i^* = \frac{3D}{2D + 1} \mu_i = \frac{D(n_i^2 + 2)}{2D + n_i^2} \mu_{i_0} \quad (73)$$

In order to eliminate E and substitute D , we make use of the relations $4\pi\mathbf{I} = (D - 1)\mathbf{E}$ and

$$\mathbf{I} = \sum_i n_i \bar{\mathbf{m}}_i \quad (74)$$

In order to eliminate the a_i 's, we introduce in place of (68) the corresponding i equations:

$$n_i \frac{4\pi}{3} a_i^3 = \zeta_i \quad (75)$$

If, as before, we assume that the molecules completely fill out the space, ζ_i represents the volume fraction occupied by the molecules of species i , and

$$\Sigma \zeta_i = 1$$

When these relations are combined with (71.1) the result is

$$D - 1 = \sum_i n_i \frac{4\pi}{3} \mu_i \mu_i^* + \sum_i \frac{3D(n_i^2 - 1)\zeta_i}{2D + n_i^2} \quad (76)$$

This is the generalization of (69) and degenerates into it when there is only one species of molecule present. As an alternative, equation (76) may be written in the form

$$\sum_i \frac{(2D + 1)(D - n_i^2)\zeta_i}{n_i^2 + 2D} = \sum_i n_i \frac{4\pi}{3} \frac{\mu_i \mu_i^*}{kT} \quad (77)$$

which is obtained from it easily by making use of the identity

$$D - 1 = \Sigma[\zeta_i(D - 1)] \quad (77.1)$$

It would, of course, be possible at this point to substitute the values of μ_i and μ_i^* in terms of D , n_i and μ_{i_0} as in the case of the one-component

system, but the result is not so instructive. Instead, we consider the limiting form assumed by (72) and (73) when $D \gg n_i^2$. In this case

$$\mu_i \rightarrow \frac{n_i^2 + 2}{3} \mu_{i_0} \quad (78)$$

and

$$\mu_i^* \rightarrow \frac{n_i^2 + 2}{2} \mu_{i_0} \quad (79)$$

At the same time, the left-hand member of (77) degenerates into

$$\sum_i D \zeta_i = D \quad (80)$$

and we obtain as a result

$$D = \frac{4\pi}{3} \sum_i n_i \frac{(n_i^2 + 2)^2}{2} \frac{\mu_{i_0}^2}{3kT} \quad (81)$$

This is to be compared with (71), of which it is a generalization.

Kirkwood's Theory

The success of the Onsager theory in accounting for the facts not only qualitatively, but to a considerable extent quantitatively also, is remarkable considering the approximations and simplifications on which it is based, notably the procedure of treating the molecule as a spherical cavity in a continuous medium having the same dielectric constant as the actual medium and replacing the actual dipole by a point dipole at the center of the cavity. It is of great interest, therefore, to find that much the same general result emerges from an alternative attack on the problem based on statistical mechanics, due to Kirkwood (1939). We shall not attempt to do more than give a rough idea of this, using a simplified argument, in order to show the extent to which the general relationships emerging from Onsager's theory are confirmed by a quite different and less arbitrary approach.

In discussing Kirkwood's theory, it is helpful to bear in mind a distinction between three different kinds of electric moments which may be associated with a molecule. First, there is the moment of the molecule in the gaseous state, where it is so far apart from other molecules that it may be regarded as completely isolated. This is the moment determined by the classical Debye procedure; we may denote it by μ_0 . In the presence of other closely adjacent molecules, especially in a polar medium, the electrical configuration of the molecule is in general altered, and its mo-

ment is different from that of the isolated molecule, although not very different. We denote this moment of the individual molecule *in solution* by \mathbf{u} , which is usually larger⁹ than \mathbf{u}_0 . Finally we must allow for the fact that any polar molecule in a liquid imposes some restraints on its near neighbors. It may exert orienting forces on them by the electrical attractions and repulsions due to its own dipole moment; it may also interfere with their rotation by steric hindrance. Thus if we specify the orientation, at a given instant, of some particular molecule which we choose arbitrarily as the center of our system, we may expect that the average orientations of its near neighbors will not be at random with respect to the central molecule. Thus the vector sum of the moment of the central molecule, plus that of all the neighbors, is in general different in magnitude, and often in direction, from \mathbf{u} . We denote this vector sum by $\bar{\mathbf{u}}$; it depends essentially on the dipole moments and the orientations of the molecules which are within two or three molecular diameters of the central molecule. Molecules which are further off are on the average oriented essentially at random with respect to the central molecule; hence their contribution to $\bar{\mathbf{u}}$ adds up to zero. The evaluation of $\bar{\mathbf{u}}$, and the formulation of its relation to the moments \mathbf{u} and \mathbf{u}_0 , is a central problem of the Kirkwood theory.

We fix our attention on a specimen consisting of a large number of molecules of the material under consideration. For the sake of simplicity in applying classical electrostatic principles, we assume the specimen to have the form of a sphere, although the final results do not depend on this form. As a consequence of their interactions the molecules of the specimen will not be oriented completely at random, even in the absence of an applied field, but there will be a certain amount of order and there will be a tendency for adjacent molecules, due to their permanent dipole moments, to be lined up in relation to one another, although the correlation between the orientations of any arbitrarily chosen pair of molecules will fall off rapidly with their mutual separation. Consider now the specimen in any *fixed* configuration of its member molecules, with its associated order. Then the moment of any particular molecule, i , in the direction of an applied field will be completely determined by the orientation of the specimen in relation to the field. We denote the permanent moment of the molecule i by the vector \mathbf{u}_i and the applied field exterior to the specimen by $\mathbf{E}_0 = E_0\mathbf{e}$, where \mathbf{e} is a unit vector in the direction \mathbf{E}_0 . The moment of the molecule in the direction of \mathbf{E}_0 is then given by

$$\mathbf{u}_i \cdot \mathbf{e} = \mu_i \cos \theta' \quad (82)$$

⁹ This will be apparent from the analysis of Onsager in the preceding section based on the concept of the reaction field, \mathbf{R} .

where θ' is the angle between \mathbf{u}_i and \mathbf{E}_0 . We emphasize that θ' is wholly determined by the orientation of the specimen in its fixed configuration. We denote by $\mathbf{M} = \Sigma \mathbf{u}_k$ the vector which gives the total moment of the specimen in this configuration as the vector sum of the moments of all

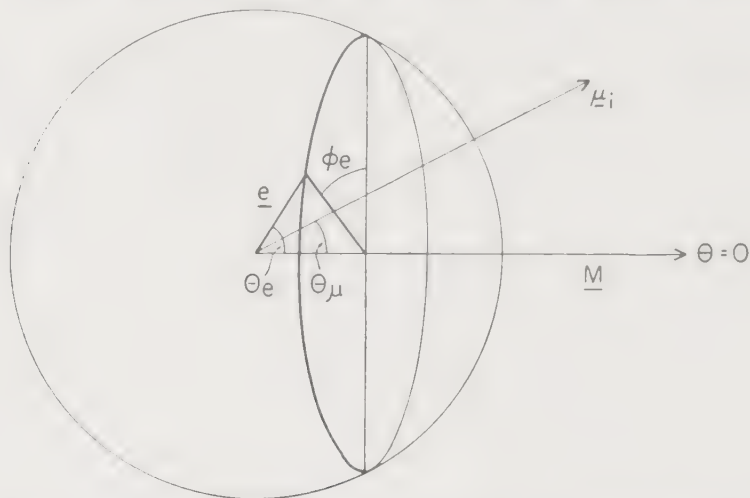


FIG. 9. The coordinate system used to describe Kirkwood's theory of dipole moments and dielectric constants.

the member molecules. Finally, we introduce a system of spherical coordinates (Fig. 9) with axis parallel to \mathbf{M} and with the plane corresponding to the azimuthal angle $\phi = 0$ parallel to \mathbf{u}_i . In terms of these coordinates, the directions of \mathbf{M} , \mathbf{u}_i , and \mathbf{e} will be as follows:

$$\begin{aligned} \mathbf{M}: \theta &= 0 \\ \mathbf{u}_i: \theta &= \theta_\mu, \phi = 0 \\ \mathbf{e}: \theta &= \theta_e, \phi = \phi_e \end{aligned} \quad (83)$$

Now let us consider the average value of $(\mathbf{u}_i \cdot \mathbf{e})$ for all orientations of the specimen *in its fixed configuration*. We may treat this specimen as a rigid body or particle whose orientation is governed by the Boltzmann distribution law. Its energy, like that of any dipole, is determined by the angle between its dipole axis and the field vector \mathbf{E}_0 and is given by $-\mathbf{E}_0 \cdot \mathbf{M} = -E_0 M \cos \theta_e$. Consequently,

$$(\mathbf{u}_i \cdot \mathbf{e})_{av} = \frac{\int_{\phi_e=0}^{2\pi} \int_{\theta_e=0}^{\pi} \mu_i \cos \theta' e^{M E_0 \cos \theta_e / kT} \sin \theta_e d\phi_e d\theta_e}{\int_{\phi_e=0}^{2\pi} \int_{\theta_e=0}^{\pi} e^{M E_0 \cos \theta_e / kT} \sin \theta_e d\phi_e d\theta_e} \quad (84)$$

We recall that θ' is the angle between \mathbf{u}_i and \mathbf{e} . In order to evaluate the expression on the right of (84) it is necessary to express θ' in terms of θ_μ , θ_e , and ϕ_e which define the orientation of \mathbf{M} and \mathbf{u}_i relative to \mathbf{e} .

This can easily be done by rewriting the well-known formula for the cosine of the angle between two straight lines in terms of spherical coordinates, and making use of (83).¹⁰ The result is

$$\cos \theta' = \cos \theta_\mu \cos \theta_e + \sin \theta_\mu \sin \theta_e \cos \phi_e \quad (85)$$

When this value of $\cos \theta'$ is introduced into (84) the integral in the numerator breaks up into two integrals, one for $\cos \theta_\mu \cos \theta_e$, and one for $\sin \theta_\mu \sin \theta_e \cos \phi_e$. The second of these vanishes, since

$$\int_0^{2\pi} \cos \phi \, d\phi = 0$$

The first remains, with the result that (84) becomes

$$[\mathbf{u}_i \cdot \mathbf{e}]_{\text{av}} = \frac{\mu_i \cos \theta_{\mu_i} \int_{\phi_e=0}^{2\pi} \int_{\theta_e=0}^{\pi} e^{ME_0 \cos \theta_e/kT} \cos \theta_e \sin \theta_e \, d\theta_e \, d\phi_e}{\int_{\phi_e=0}^{2\pi} \int_{\theta_e=0}^{\pi} e^{ME_0 \cos \theta_e/kT} \sin \theta_e \, d\theta_e \, d\phi_e}$$

This is the same, except for differences involving μ , M , and $\cos \theta_{\mu_i}$, as the expression for $\overline{\cos \theta}$ in the original Debye theory, and may be evaluated in the same way, with the result

$$[\mathbf{u}_i \cdot \mathbf{e}]_{\text{av}} = \mu_i \cos \theta_{\mu_i} \left(\coth a - \frac{1}{a} \right) \quad (86)$$

where $a = ME_0/kT$. Here, therefore, we encounter the Langevin function once again, and for small values of the argument ME_0/kT we have

$$[\mathbf{u}_i \cdot \mathbf{e}]_{\text{av}} = \mu_i \cos \theta_{\mu_i} \frac{ME_0}{3kT} = \mathbf{M} \cdot \mathbf{u}_i \frac{E_0}{3kT} \quad (87)$$

It should be realized that $[\mathbf{u}_i \cdot \mathbf{e}]_{\text{av}}$, is the mean value of the moment of the i th molecule in the direction of the applied field *for all possible orientations of the specimen in a given fixed configuration*. Actually what we really want in order to obtain an expression involving the dielectric constant is the mean value of \mathbf{u}_i for *any molecule* in the direction of the field for *all configurations* as well as *all orientations* of the specimen. This may be obtained as follows. It will be recalled that θ_{μ_i} is the angle between

¹⁰ If α , β , γ denote the direction cosines of any vector with respect to the three Cartesian axes x , y , z , of which x is parallel to the polar axis $\theta = 0$, the plane $\phi = 0$ being defined by the x axis and the positive y axis, then $\alpha = \cos \theta$, $\beta = \sin \theta \cos \phi$, $\gamma = \sin \theta \sin \phi$. Thus in Fig. 9, the direction cosines of \mathbf{u}_i , with respect to the x , y , z axes are $l = \cos \theta_\mu$, $m = \sin \theta_\mu$, $n = 0$, respectively. The direction cosines of \mathbf{e} are $l' = \cos \theta_e$, $m' = \sin \theta_e \cos \phi_e$, $n' = \sin \theta_e \sin \phi_e$. We then apply the general formula for the cosine of the angle between two vectors with specified direction cosines: $\cos \theta' = ll' + mm' + nn'$, which immediately gives (85).

\mathbf{u}_i and the total moment, \mathbf{M} , of the specimen in a fixed configuration. When we average $M \cos \theta_{\mu_i}$ over all configurations of the specimen, we are, therefore, in reality reckoning the mean moment of the specimen in the direction of \mathbf{u}_i . The mean value of $M_{\mu_i} \cos \theta_{\mu_i}$, averaged over all configurations may, therefore, be expressed as the scalar product $\mathbf{u}_i \cdot \bar{\mathbf{M}}_i$, where $\bar{\mathbf{M}}_i$ is interpreted as the mean moment of the specimen produced by its i th molecule. When we consider all configurations of the specimen, however, the distinction between the molecules disappears—for they are all of the same kind—and we may omit the subscript i . The final result is consequently

$$(\mathbf{u} \cdot \mathbf{e})_{\text{av}} = \frac{(\bar{\mathbf{M}} \cdot \mathbf{u})E_0}{3kT} \quad (88)$$

Here $(\mathbf{u} \cdot \mathbf{e})_{\text{av}}$ is simply the mean moment of any molecule in the direction of the applied field, without further qualification; $\bar{\mathbf{M}}$ is the mean total moment of the specimen produced by any molecule within it—it will be the same for all.

The result given in (88) is expressed in terms of the field \mathbf{E}_0 external to the specimen. This arises from our procedure of treating the specimen as a rigid particle in empty space and applying the Boltzmann distribution law to it. Actually what we are interested in is the mean value $(\mathbf{u} \cdot \mathbf{e})_{\text{av}}$ in terms of the field \mathbf{E} prevailing in the *interior* of the specimen. It is easy to make the transition from \mathbf{E}_0 to \mathbf{E} . Since the specimen is spherical, all that we need in order to do this is the expression for the field \mathbf{E} inside a sphere of dielectric constant D when the field outside in empty space is \mathbf{E}_0 . This may be obtained directly from the more general expression (37) derived in an earlier section by replacing D_1 by D and D_2 by 1. The result is

$$\mathbf{E} = \frac{3\mathbf{E}_0}{(D + 2)} \quad (89)$$

By making use of this, we obtain at once the desired relation

$$(\mathbf{u} \cdot \mathbf{e})_{\text{av}} = \frac{(D + 2)}{9kT} (\mathbf{u} \cdot \bar{\mathbf{M}})E \quad (90)$$

The possibility of this simple transition is the reason for taking the specimen as spherical.

It may be repeated and should be emphasized that $\bar{\mathbf{M}}$ is to be interpreted as the average *total* moment induced in a spherical specimen by any one of its member molecules maintained in a fixed position.¹¹ By applying classical electrostatic considerations to the specimen, involving

¹¹ Including, of course, the moment of the molecule itself.

solutions of Laplace's equation both within it and without, it is possible to show that $\bar{\mathbf{M}}$ is a finite quantity, independent of the size of the specimen, as is intuitively demanded, and, furthermore that it may be conveniently decomposed into two parts, which bear a constant ratio to one another. One of these, which we denote by $\bar{\mathbf{M}}_1$, is distributed uniformly throughout the specimen and may be interpreted as arising from the reaction field at its outer boundary produced by the molecule in question. The other, which we denote by $\bar{\mathbf{u}}$, is in contrast concentrated round the same molecule. By increasing the radius of the specimen to infinity, it is possible to reduce the portion of $\bar{\mathbf{M}}_1$ contained in any finite region, or indeed in any infinite region of lower order than the specimen, to zero. This provides a basis for the definition of $\bar{\mathbf{u}}$, for $\bar{\mathbf{u}}$ may then be defined as the mean moment of a sphere with its center at the molecule in question, whose radius, though infinite, is of a lower order than that of the specimen. In practice, the infinite sphere may be replaced by a small sphere which, however, must be large enough so that at its surface, from the point of view of an observer at the center, the dielectric constant may be fairly taken as the same as the macroscopic dielectric constant. The radius of such a sphere will actually be of molecular dimensions. It can be shown (see Kirkwood, 1939) that $\bar{\mathbf{u}}$ is related to $\bar{\mathbf{M}}$ by

$$\bar{\mathbf{M}} = \left[\frac{3}{(D+2)} \frac{3D}{(2D+1)} \right] \bar{\mathbf{u}} \quad (91)$$

Since, for any value of $D > 1$, $[3/(D+2)][3D/(2D+1)]$ is always less than 1, it is evident that $\bar{\mathbf{u}}$ and $\bar{\mathbf{M}}_1$ must partially oppose one another.

We are now in a position to express the total polarizability in terms of the dielectric constant. For this purpose, we have recourse to the familiar relation $(D-1)E = 4\pi n\bar{m}$, where n is the number of molecules per unit volume and \bar{m} is the total mean moment of a molecule in the direction of the field, taking account both of the induced moment and that due to the orientation of the molecule as a permanent dipole. Since the induced moment is given as in earlier discussions by $\alpha_0 F$, this may be written as

$$(D-1)E = 4\pi n \left[\alpha_0 \frac{F}{E} + \frac{(D+2)\bar{\mathbf{u}} \cdot \bar{\mathbf{M}}}{9kT} \right] E \quad (92)$$

or, after substitution of the value of $\bar{\mathbf{M}}$ given by (91),

$$(D-1) = 4\pi n \left[\alpha_0 \frac{F}{E} + \frac{3D}{3kT(2D+1)} \bar{\mathbf{u}} \cdot \bar{\mathbf{u}} \right] \quad (93)$$

We are primarily interested in the case of strongly polar media, in which

the induced moment, $\alpha_0 F$, is a small fraction of the total. It is not of great importance, therefore, to the over-all result just what value we take for F/E . It is convenient to introduce the value used in the Onsager theory, namely

$$\frac{F}{E} = \frac{3D}{(2D + 1)} \quad (94)$$

though this ties us, to a very minor degree, to that theory. The result is

$$\frac{(D - 1)(2D + 1)}{9D} = \frac{4}{3} \pi n \left(\alpha_0 + \frac{\mathbf{u} \cdot \bar{\mathbf{u}}}{3kT} \right) \quad (95)$$

This is the fundamental equation of the Kirkwood theory for a one-component system.

Except for the difference between $\mathbf{u} \cdot \bar{\mathbf{u}}$ and μ_0^2 , the right-hand member of (95) is the same as the volume polarizability, p , of the Debye theory. By analogy it is, therefore, defined as the volume polarizability in the Kirkwood theory:

$$p = \frac{4}{3} \pi n \left(\alpha_0 + \frac{\mathbf{u} \cdot \bar{\mathbf{u}}}{3kT} \right) \quad (96)$$

Similarly $\frac{4}{3} \pi \left(\alpha_0 + \frac{\mathbf{u} \cdot \bar{\mathbf{u}}}{3kT} \right)$ is defined as the molecular polarizability, and the corresponding molar quantity is defined as the molar polarizability, both definitions being, of course, like (96), peculiar to the Kirkwood theory.

$$P = \frac{4}{3} \pi N \left(\alpha_0 + \frac{\mathbf{u} \cdot \bar{\mathbf{u}}}{3kT} \right) \quad (97)$$

Actually the two quantities μ_0^2 and $\mathbf{u} \cdot \bar{\mathbf{u}}$ should not be *very* different, though it is to be expected that $\mathbf{u} \cdot \bar{\mathbf{u}}$ will always be greater than μ_0^2 , for \mathbf{u} itself should be somewhat greater than \mathbf{u}_0 , owing to polarization by the reaction field, as analyzed by Onsager. Moreover, $\bar{\mathbf{u}}$, being the total moment of the central molecule and its neighbors in a spherical region within which the dielectric constant differs significantly from its macroscopic value, must be greater still. Theoretically $\bar{\mathbf{u}}$ and \mathbf{u} need not be parallel to one another except in the case of molecules having a threefold or higher axis of symmetry.

The left-hand member of (95) is closely similar to that of (40) derived by the simple expedient of treating the molecule as a spherical cavity of low dielectric constant containing any sort of dipole at an unspecified position within it. In fact, the two become identical if we set D_1 , the dielectric constant of the cavity in (40), equal to unity. The right-hand

member of (40) is, of course, the same as the volume polarizability of the Debye theory. The close similarity of equations (40) and (95) is striking, considering the wholly different way in which they were derived. As we have said, the difference between μ_0^2 and $\mathbf{u} \cdot \bar{\mathbf{u}}$ will not be very great—certainly the two quantities should be in the same range of magnitude, as we shall see in the next section. This comparison suggests, therefore, that the principal source of difficulty with the Debye theory as applied to polar media arises from failure to take account of the dielectric properties of the molecules rather than from neglect of their interactions.

The connection between (95) and (70), the latter of which embodies the Onsager theory for a one-component medium, is not quite so clear, though it is no less real. It may best be seen when we compare the limiting forms assumed by these equations for strongly polar media where $D \gg n^2$ and therefore $\mu^2/3kT \gg \alpha_0$. These may be written as

$$\frac{2D}{(n^2 + 2)^2} \cong \frac{4\pi n \mu_0^2}{9kT} \quad (\text{Onsager}) \quad (98)$$

$$\frac{2D}{9} \cong \frac{4\pi n}{3kT} \left(\alpha_0 + \frac{\mathbf{u} \cdot \bar{\mathbf{u}}}{3kT} \right) \cong 4\pi n \frac{\mathbf{u} \cdot \bar{\mathbf{u}}}{9kT} \quad (\text{Kirkwood}) \quad (99)$$

Taken as they stand, (98) and (99) imply that

$$\mathbf{u} \cdot \bar{\mathbf{u}} = \frac{(n^2 + 2)^2}{9} \mu_0^2$$

with n the internal index of refraction of a molecule. In the case of water, if we assign to n the macroscopic value of the index of refraction (1.33) we find

$$\mathbf{u} \cdot \bar{\mathbf{u}} = 1.56\mu_0^2$$

Thus \mathbf{u}_0 , the quantity in which we are generally interested, may be expected to be less than $\mathbf{u} \cdot \bar{\mathbf{u}}$ by something like 25%. This would mean that, if we used the simple modification of the Debye theory first introduced (40), our calculated moments would be expected to be high by about that amount. All these theories are somewhat arbitrary and artificial, however, and it is necessary to proceed carefully and empirically.¹²

¹² It is possible to show that if, in the Kirkwood treatment, the radius of the sphere surrounding the central molecule in which the dielectric constant differs from its macroscopic value were reduced to the radius of a molecule, the theory would become equivalent to the Onsager theory. In that case, of course, $\bar{\mathbf{u}}$ would become the same as \mathbf{u} . We have seen that the Kirkwood theory gives the same result as the simpler treatment which leads to (40) except for the difference between μ_0^2 and $\mathbf{u} \cdot \bar{\mathbf{u}}$. It follows, therefore, that the effect of taking account of the reaction field in the Onsager theory is represented by the difference between μ_0^2 and μ^2 .

In the case of solutions or, more generally, systems of several components, equations (90) and (91) hold for each kind of molecule present. Therefore, if we carry through the calculations, equation (95) becomes

$$\frac{(D-1)(2D+1)}{9D} = \sum \frac{4\pi}{3} n_k \left(\alpha_{0k} + \frac{\mathbf{u}_k \cdot \bar{\mathbf{u}}_k}{3kT} \right) \quad (100)$$

which shows that the Kirkwood polarizabilities are additive. This equation is the full expression of the Kirkwood theory for the general case of systems of several components. The situation here is much simpler than in the Onsager theory where no explicit equation corresponding to (100) is easily given (see equation 78).

In the case of highly polar systems, equation (100) passes over into the limiting form, which is a generalization of (99).

$$\frac{2D}{9} \cong \frac{4}{3} \pi \sum_k n_k \left(\alpha_{0k} + \frac{\mathbf{u}_k \cdot \bar{\mathbf{u}}_k}{3kT} \right) \cong \frac{4}{3} \pi \sum_k n_k \left(\frac{\mathbf{u}_k \cdot \bar{\mathbf{u}}_k}{3kT} \right) \quad (101)$$

The corresponding equation of the Onsager theory may be written in the form

$$2D = \frac{4}{3} \pi \sum n_k \left[\frac{(n_k^2 + 2)^2}{3kT} \right] \mu_k^2 \quad (102)$$

In closing this section, it is appropriate to call attention to another, more complete and more abstract approach to the problem of the dielectric constant, based on concepts of quantum theory, which has been given by van Vleck (1937). Unfortunately, the equations, though rigorous, cannot be solved to give much information about actual situations, but they do show (and this is significant) that the generalized Clausius-Mossotti relation must break down and the "4 π I/3 catastrophe" thereby be avoided in highly polar media. This is in accord, as far as it goes, with the more limited treatments of Onsager and Kirkwood and all that we know from experience.

Polar Liquids and Dipolar Ions

We have already referred to the empirical relation obtained by correlating the dielectric constant of a large number of polar liquids with the total volume polarizability calculated according to the Debye definition (16) from known values of the index of refraction, density, and electric moment as measured under ideal conditions (i.e., in the vapor state or in dilute solution in nonpolar solvents). Apart from certain exceptional cases, involving molecules like water and the alcohols, which are commonly thought to be strongly associated owing to hydrogen bonding and which

exhibit various other abnormalities of behavior, the data show that for liquids of dielectric constant greater than 5 the volume polarizability is very roughly linear in the dielectric constant and is given empirically by

$$p = \frac{D + 1}{8.5} \quad (103)$$

The numerical constant 8.5 is subject to considerable uncertainty, however, owing to scatter of the data. For large values of D , characteristic of strongly polar solutions, (103) passes over into the limiting form

$$p = \frac{D}{8.5} \quad (104)$$

This corresponds to the limiting equation (99) of the Kirkwood theory, which may be written

$$p = \frac{2}{9} D = \frac{D}{4.5} \quad (105)$$

where p is the volume polarizability defined in terms of $\mathbf{u} \cdot \bar{\mathbf{u}}$. In Fig. 10, we have plotted for comparison graphs of the three equations: (103), (95), which is the full equation of the Kirkwood theory, and (16), which embodies the classical Debye theory. It will be noted that (95) rapidly approaches the asymptote $p = (2D - 1)/9$ which cuts the abscissa axis at $D = 1/2$ just as the empirical equation (103) extrapolates back to $D = -1$ at $p = 0$. Both points ($D = 1/2$ and $D = -1$) have, of course, no physical meaning, and the corresponding equations are of significance only for values of D greater than about 5.

The empirical relation (103), (or 104) will be seen to accord well with the predictions of the Kirkwood theory when it is recalled that $\mathbf{u} \cdot \bar{\mathbf{u}}$ is always greater than μ_0^2 . Indeed (103) provides the basis for a clearer appreciation of the significance of $\mathbf{u} \cdot \bar{\mathbf{u}}$, the fundamental quantity of the Kirkwood theory. As a first approximation we may assume the ratio of μ_0^2 to $\mathbf{u} \cdot \bar{\mathbf{u}}$ to be constant from one substance to another, apart from a few highly unusual substances, such as water. Then if we neglect α_0 in comparison with μ_0^2 (and, *a fortiori*, $\mathbf{u} \cdot \bar{\mathbf{u}}$), which is justifiable in the case of strongly polar liquids, division of (104) and (105) gives, for the absolute value of μ_0 ,

$$\mu_0 = \sqrt{\frac{4.5}{8.5} \mathbf{u} \cdot \bar{\mathbf{u}}} = 0.73 \sqrt{\mathbf{u} \cdot \bar{\mathbf{u}}} \quad (106)$$

This means that the moment of a substance in the gas phase or in dilute solution in a nonpolar solvent should be about three-fourths the value of $\sqrt{\mathbf{u} \cdot \bar{\mathbf{u}}}$ calculated by the Kirkwood theory, or, what amounts to the same

thing, the value of the moment calculated by the first proposed modification of the Debye theory, from data on strongly polar media. This agrees with deductions drawn in the last section in the course of a comparison of the Onsager and Kirkwood theories. It is an important conclusion which will serve as a guide in the future interpretation of results on the dielectric properties of polar liquids and solutions.

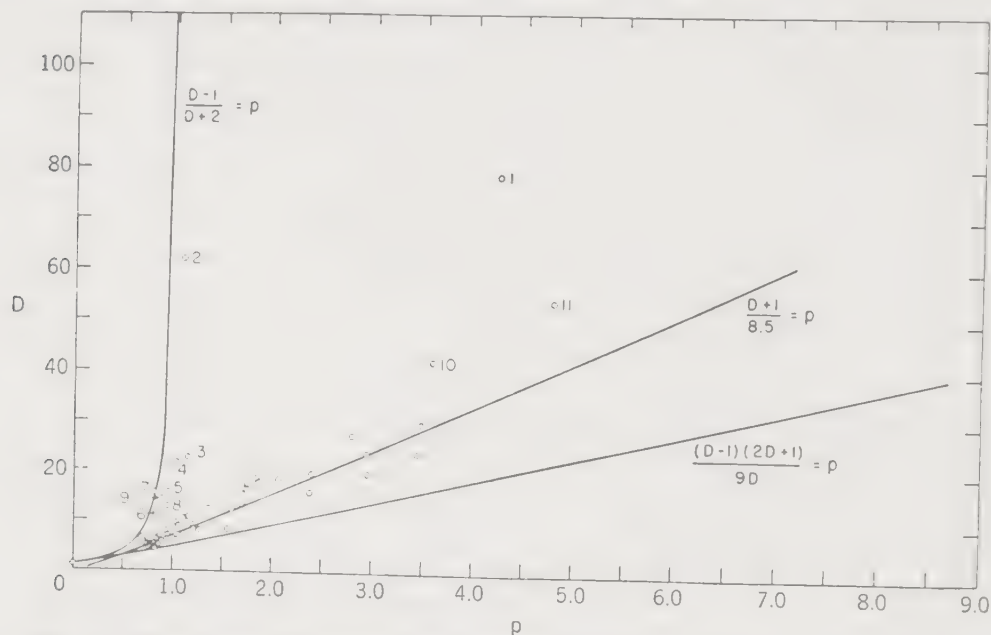


FIG. 10. The relation between the dielectric constants of certain liquids and the polarization (p) of the same substances in the vapor state. The three solid lines represent the calculated relations between D and p according to the equations of Debye, Wyman and Kirkwood. Each circle shows the experimental values of D and p for a particular molecule. The numbered circles are as follows: 1, water; 2, formic acid; 3, epichlorohydrin; 4, propyl alcohol; 5, *s*-butyl alcohol; 6, *t*-butyl alcohol; 7, amyl alcohol; 8, benzyl alcohol; 9, paraldehyde; 10, furfuraldehyde; 11, nitrosodimethylamine. For identification of the other substances represented in the figure, see Wyman (1936a).

Of course, if we are interested in the highest accuracy, we must expect the ratio of $\mu_0^2/\mathbf{u} \cdot \bar{\mathbf{u}}$ to vary somewhat from one medium to another, for $\mathbf{u} \cdot \bar{\mathbf{u}}$, involving the interaction of a molecule and its neighbors, will be dependent on the special circumstances of each case. In terms of the Onsager theory (equation 70 or 71) this variation may be attributed to differences in the effective internal index of refraction of the molecules. It will be seen from (71) that the constant 8.5 of equation (104) corresponds to

$$(n^2 + 2)^2 = 17$$

or $n = 1.45$. But in general it is preferable to analyze the data on the

basis of the Kirkwood theory in terms of $\mathbf{u} \cdot \bar{\mathbf{u}}$, without raising the issue of the exact relation between this and μ_0^2 .

The quantity $\mathbf{u} \cdot \bar{\mathbf{u}}$ represents a statistical concept. In order to estimate it, Kirkwood has proposed an approximation in which it is given by

$$\mathbf{u} \cdot \bar{\mathbf{u}} = \mu_0^2(1 + z \overline{\cos \gamma}) \quad (107)$$

Here z is the number of neighbor molecules in the region around a given molecule within which, from the point of view of an observer stationed at the molecule, the effective dielectric constant deviates significantly from its macroscopic value. On the other hand $\overline{\cos \gamma}$ is the average value of the cosine of the angle between the dipole moments of any pair of neighbor molecules in the region. Kirkwood has used this relation to calculate the value of $\mathbf{u} \cdot \bar{\mathbf{u}}$ for water on the basis of the extensive information on the structure of that liquid. (See Chapter 2.) From the value so obtained he arrives, by means of equation (99), at a figure for the dielectric constant which agrees quite well with the experimental value at 25°. It may be remarked that the calculation is very sensitive to small differences in the exact assumptions relating to the structure of the liquid. Water is a rather special case. Kirkwood assumes that each water molecule is coordinated with four nearest neighbors ($z = 4$) according to the arrangement portrayed in Chapter 2, Fig. 3. Beyond the nearest neighbors, he takes the water as a continuous medium, characterized by its macroscopic dielectric constant. His results are as follows:

THE DIELECTRIC CONSTANT OF WATER AT 25°
(Kirkwood's calculations)

Assumed O ^H _H bond angle	μ_0	$\overline{\cos \gamma}$	$\mathbf{u} \cdot \bar{\mathbf{u}}/\mu_0^2$	D_{calc}	D_{obs}
90°	1.88	0.50	4.35	82	78.54
100°	1.88	0.41	3.55	67	
109° 28'	1.88	0.33	2.91	55	

DIELECTRIC CONSTANTS AND DIPOLE MOMENTS OF DIPOLAR IONS IN SOLUTION

We turn now to a consideration of the dielectric properties of dipolar ions in solution. This is a large subject, and the study of it has been revealing not only in respect to the electrical properties of the dipolar ions themselves, but also in relation to the theory of the dielectric constant. Indeed to a considerable degree the theoretical considerations given in preceding sections were provoked by the extensive and suggestive body of data on the dielectric behavior of dipolar ions, notably the amino acids

and their derivatives, in solution. It is no accident, but indeed accords with their own intensely polar structure, that these substances are virtually insoluble in organic, and particularly in nonpolar, solvents, but that they dissolve readily in water and salt solutions. For this reason most of what we know about their dielectric behavior relates to aqueous solutions. We turn first to the case of glycine.

Glycine has been studied in water solution up to a concentration of over 2.5 moles per liter, which is not far from the limit of its solubility. The dielectric constant increases rapidly and linearly with the concentration of the amino acid, reaching a value of about 135 at a concentration of 2.5 moles per liter at 25°, and is given (at 25°) by

$$D = 78.54 + 22.58C$$

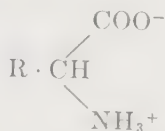
where C is concentration in moles per liter.

The mean departure (regardless of sign) of the experimental points from the values calculated from this linear relation is, in the most reliable experiment, no more than 0.3% and is unsystematic. The great increase of the dielectric constant with concentration reflects the extremely large moment of glycine as a dipolar ion, and the linearity of the relationship represents the proportionality between dielectric constant and polarizability characteristic of strongly polar media.

Essentially the same behavior is shown by other α -amino acids. Not only do they all display the same linear increase of dielectric constant with concentration, but the numerical value of the slope of the curve, dD/dC , is very nearly the same for all. Thus, for example, in the case of α -aminobutyric acid in water at 25° the dielectric constant is given by

$$D = 78.54 + 23.53C$$

up to a concentration of 2.04 moles per liter, which is about the limit of its solubility, and the mean departure, regardless of sign, of the experimental points from this formula is less than 0.2%. Since all the α -amino acids have virtually the same moment, the significant dipolar ionic configuration



being common to all, this means that the quantity dD/dC is a direct expression of the molar polarizability of the amino acid. This inescapable conclusion is borne out by the fact that other amino acids and dipolar ions of a given type, e.g., the β -amino acids, all show the same exact linearity between dielectric constant and concentration and are all char-

acterized by another common value of dD/dC , which is greater, the greater the separation of the charge groups in the amino acid. Moreover, for a given amino acid the effect on the dielectric constant at a given concentration is essentially independent of the nature of the solvent and its dielectric constant, as shown by the data given below in Table III on glycine.

The quantity dD/dC , which shows itself to be such a direct reflection of the polarizability of the molecules, is known as the molar dielectric increment. It is usually designated in the literature by the small Greek letter δ . The significance of the dielectric increment and the reason for its close relation to the polarizability of the solute molecules is clear from the theoretical considerations of preceding sections and will no doubt have already suggested itself to the reader. If we multiply both numerator and denominator of Kirkwood's equation (101) by Avogadro's number, we obtain

$$\frac{2}{9}D = \sum_k c_k P_k \quad (108)$$

in which the c 's denote concentrations in moles per unit volume and the P 's denote total molar polarizabilities defined in terms of $\mathbf{u} \cdot \mathbf{\bar{u}}$ (see equation 97). For a two-component solution, this becomes

$$\frac{2}{9}D = c_1 P_1 + c_2 P_2 \quad (109)$$

It should be realized that c_1 and c_2 are not independent of one another but are connected by the relation

$$c_1 V_1 + c_2 V_2 = 1 \quad (110)$$

where the V 's are the molar volumes, partial or apparent. If we assume that the V 's are independent of the concentration, thereby identifying the apparent and partial volumes, and also treat the P 's as constant, it follows from these two equations that

$$\frac{2}{9} \frac{dD}{dc_2} = P_2 + P_1 \frac{dc_1}{dc_2} = P_2 - \frac{P_1 V_2}{V_1} \quad (111)$$

If we associate subscript 2 with the solute and subscript 1 with the solvent, dD/dc_2 becomes the same as the dielectric increment except for a small complication involving units. The P 's of course have the dimensions of volume, and it is universal practice to express them in cubic centimeters. This means that c_2 in (111) represents moles per cubic centimeter. The dielectric increment, on the other hand, is always expressed in terms of concentration in moles per liter (C), which is 1000

times as great. By taking account of this, and introducing δ as the symbol for the dielectric increment into (111), we obtain

$$\frac{2 \times 1000\delta}{9} = P_2 - \frac{P_1 V_2}{V_1}$$

In this expression P_1/V_1 is simply the volume polarizability of the pure solvent, which is given by $p_1 = (\frac{2}{9})D_1$, D_1 being the dielectric constant of the solvent. Consequently, as a final result we have

$$P_2 = \frac{2}{9}(1000\delta + D_1 V_2) \quad (112)$$

The same procedure applied to the Onsager equation (102) yields the slightly more involved expression

$$(P_2 - P_{20}) = \frac{2}{(n_1^2 + 2)^2} (1000\delta + D_1 V_2) \quad (113)$$

It should be borne in mind that P_2 in (112) is defined as

$$P_2 = \frac{4}{3} \pi N \left(\alpha_{20} + \frac{\mathbf{u}_2 \cdot \bar{\mathbf{u}}_2}{3kT} \right)$$

and $P_2 - P_{20}$ in (113) is

$$P_2 - P_{20} = \frac{4}{9} \pi N \mu_{20}^2$$

where μ_{20} is the ideal moment of the molecule in the vapor phase. For our purposes it is better to use (112), expressing our results in terms of $\mathbf{u} \cdot \bar{\mathbf{u}}$, and from now on we shall do so. Equations (112) and (113), though derived for a two-component system, are, of course, applicable to a system in which the solvent is itself a solution, provided its composition is held constant. This is self-evident but could, if it were desired, be justified by the application of the procedure given above to this more general case.

For such highly polar substances as dipolar ions, the dominating term on the right in either (112) or (113) is 1000δ . Thus in the case of glycine in aqueous solution $D_1 V_2$ is only about 3500, as compared with $1000\delta = 22,580$, as we shall see below. This helps explain the very precisely linear increase of dielectric constant with concentration exhibited by these substances in solution, for changes in the molar volumes can only affect the term $D_1 V_2$. (Actually, the partial molar volumes are found to be only slightly dependent on concentration.) It also explains the nearly identical values of the dielectric increments of amino acids of a given type, which, of course, have the same moments. Likewise, it accounts for the fact that the dielectric increment of a given amino acid

is extremely insensitive to changes in the dielectric constant of the solvent, for a change of solvent can directly affect the dielectric increment only through the term D_1V_2 . In short, equations (112) and (113) immediately render intelligible the whole remarkably simple and revealing behavior of the dielectric constant of dipolar ions in solution.

We now apply (112) to calculate the value of $\mathbf{u} \cdot \bar{\mathbf{u}}$ for glycine in water on the basis of the data given above. The first step is to obtain the total polarizability, P . If we take the molar volume, V , as 43.5 cc, we obtain from the value of $\delta = 22.58$

$$P = 5790 \text{ cc}$$

To pass from this to $\mathbf{u} \cdot \bar{\mathbf{u}}$ we may employ an equation analogous to (25), writing it as

$$\sqrt{\mathbf{u} \cdot \bar{\mathbf{u}}} = 0.0127 \sqrt{(P - P_0)T} \text{ Debye units} \quad (114)$$

This is justified by the fact that the definition of P in the Kirkwood theory is the exact counterpart of that in the Debye theory, $\mathbf{u} \cdot \bar{\mathbf{u}}$ replacing μ_0^2 . For P_0 we may use the value of the optical polarizability of glycine ethyl ester, which is 25.73 cc, though this is really not large enough to warrant consideration. The result is then

$$\sqrt{\mathbf{u} \cdot \bar{\mathbf{u}}} = 16.7 \text{ Debye units}$$

As we have said, the value μ_0 , the ideal moment which the molecule would have in the absence of its neighbors, may be presumed to be about 75%

TABLE III
POLARIZABILITIES AND MOMENTS OF GLYCINE AT 25°
(Molecular weight 75, $P_0 = 26$ cc)

Solvent	D^*	Concentration range studied	δ	Partial molal volume	P	$\sqrt{\mathbf{u} \cdot \bar{\mathbf{u}}}$
60% ethyl alcohol	47.88	0-0.133	20.4	42.6	5,000	15.5
40% ethyl alcohol	59.69	0-0.470	21.7	45.7	5,340	16.0
20% ethyl alcohol	69.96	0-1.27	22.55	44.6	5,700	16.5
Water	78.54	0-2.5	22.58	44.3	5,790	16.7
2.5 <i>M</i> urea	84.35	0-2.17	22.3	46.1	5,800	16.7
5.0 <i>M</i> urea	90.60	0-2.45	22.6	46.5	5,960	16.9
0.590 <i>M</i> α -aminobutyric acid	92.30	0-2.40	21.9	45.2	5,790	16.7
1.198 <i>M</i> α -aminobutyric acid	106.75	0-2.30	21.0	46.4	5,750	16.6
1.826 <i>M</i> α -aminobutyric acid	121.5	0-2.20	20.6	46.6	5,840	16.7

* D denotes the dielectric constant of the solvent, without added glycine.
For references, see J. Wyman, *Chem. Revs.* **19**, 213 (1936).

TABLE IV

POLARIZABILITIES AND MOMENTS OF REPRESENTATIVE DIPOLAR IONS
(Unless otherwise specified all results are from data on aqueous solutions at 25°. In calculating the moments optical polarizabilities have been neglected.)

Substance	δ	Partial molal volume	P	$\sqrt{\mu \cdot \bar{\mu}}$ Debye units
Glycine	22.6	43.5	5,790	16.7
α -Alanine	23.3	60.6	6,200	17.2
α -Aminobutyric acid	23.5	76.3	6,560	17.7
α -Aminovaleric acid	22.6	92.7	6,650	17.9
β -Alanine	34.8	58.9	8,760	20.6
β -Aminobutyric acid	34.2	76.4	8,950	20.8
γ -Aminobutyric acid	51.0	(76.4)	12,600	24.6
γ -Aminovaleric acid	54.8	90.0	13,700	25.6
δ -Aminovaleric acid	63	(90)	15,700	27.5
ϵ -Aminocaproic acid	75	105	18,500	29.8
ζ -Aminoheptylic acid	87	(116)	21,400	32.1
L-Glutamic acid	26	(90)	7,350	18.8
L-Arginine	62	(78)	15,200	27.1
Creatine	32.2	87	8,660	20.4
Acetylhistidine	62	134	16,100	27.8
Glycine dipeptide	71	77	17,100	28.6
Glycine tripeptide	120	115	28,700	37.2
Glycine tetrapeptide	159	151	38,000	42.6
Glycine pentapeptide	215	190	51,100	49.5
Glycine hexapeptide	234	(231)	56,100	52.0
Glycine heptapeptide in 5.14 <i>M</i> urea, $D = 91.36$	290	217	68,900	57.8
Glycylphenylalanine	70.4	155	18,400	29.7
L-Leucylglycylglycine	120.4	178	29,800	37.8
Lysylglutamic acid	345	173	79,600	61.9
ϵ - ϵ' -Diamino-di(α -thio- <i>n</i> -caproic acid)	131	227	33,100	39.8
ϵ - ϵ' -Diguanido-di(α -thio- <i>n</i> -caproic acid)	151	(300)	33,100	43.1
Cystinyldiglycine	139	194	34,300	40.6
Glycine betaine	18.2	98	5,750	16.6
ξ -Aminopentadecylic acid betaine at 70°, $D = 64$	220	(250)	52,500	54.3
π -Aminoheptadecylic acid betaine at 80°, $D = 61$	190	(280)	46,000	51.3
<i>o</i> -Benzbetaine	18.7	147	6,710	18.0
<i>m</i> -Benzbetaine	48.4	145	13,300	25.3
<i>p</i> -Benzbetaine	72.4	141	18,500	29.8
Horse hemoglobin, M.W. 67,000	22,100	49,100	5.8×10^6	*
Serum mercaptalbumin, M.W. 69,000	60,000	50,500	12.1×10^6	*
β -Lactoglobulin, 25°, M.W. 40,000	60,000	30,000	13.9×10^6	*
Egg albumin, M.W. 44,000	4,400	33,000	1.5×10^6	*

For references, see J. Wyman (1936b); J. L. Oncley, *Chem. Revs.* **30**, 433 (1942); and E. J. Cohn and J. T. Edsall (1943), Chapter 22. The figure for serum mercapt-

of this, or $\mu_0 = 12.5$. These figures for $\sqrt{\mathbf{u} \cdot \bar{\mathbf{u}}}$ and μ_0 may both be compared with the estimate $\mu = 15$ derived in Chapter 5 from the effect of ionic strength on solubility and with the value $\mu_0 = 14.0$ which may be estimated from structural considerations. It is of interest to see that the figures 16.7 for $\sqrt{\mathbf{u} \cdot \bar{\mathbf{u}}}$ and 15 for \mathbf{u} imply that $\bar{\mathbf{u}} = 19.3$, if $\bar{\mathbf{u}}$ and \mathbf{u} are assumed to be parallel.

Table III summarizes the results of similar calculations applied to the large body of data on glycine in a variety of mixed solvents covering a range of dielectric constant from 48 to 122 in all of which the dielectric increment remains about the same, as mentioned earlier. The striking feature of the results is the essential constancy of the polarizability and of $\sqrt{\mathbf{u} \cdot \bar{\mathbf{u}}}$, which is even greater than that of the dielectric increment, variation in the latter tending to be compensated by changes in D_2V_1 , except in the case of 40 and 60% ethanol. The drop of polarizability in these two solvents—the solvents of lowest dielectric constant among those employed—is probably real and represents a difference in the interaction between the amino acid and its neighbor molecule rather than a change in the intrinsic moment of the amino acid. Even in 60% ethanol the concentration ratio of the dipolar ionic to the undissociated form of glycine



is very large indeed (see Chapters 8 and 9, and the discussion in Chapter 4 of Cohn and Edsall, 1943). It should be realized that in solution the value of $\mathbf{u} \cdot \bar{\mathbf{u}}$ is a quantity which does not depend exclusively on the properties of the solute molecules but involves also their interaction with the molecules of the solvent.

Table IV gives a survey of data of the dielectric increments, polarizabilities, and moments of a variety of representative dipolar ions in aqueous solution, all calculated on the basis of equation (112). The immediately striking feature of this table is the size of the polarizabilities and moments. Of course, it should be understood that these are calculated in accordance with the Kirkwood definition and that the ideal moments (μ_0) are certainly somewhat smaller than the values of $\sqrt{\mathbf{u} \cdot \bar{\mathbf{u}}}$ given in the table, but, even allowing for this, the magnitude of the results goes

albumin is from H. M. Dintzis, thesis, Harvard University (1952). We are indebted to Prof. J. L. Oncley for making this thesis available to us, and for valuable discussions.

* Values for proteins could be calculated from the given P values by equation (114), but we have refrained from recording them, since the interpretation of the observed dielectric increments in terms of permanent electric moments is uncertain for molecules containing many acid and basic groups, which take up and give off protons rapidly near the isoelectric point. See the discussion of dielectric dispersion in Volume II.

far to satisfy human cravings for large numbers. The near identity of the moments of dipolar ions of a given type is impressive, although it is curious that the effect of the second term in equation (112) is actually to increase slightly, rather than to diminish, the differences between the amino acids (compare, for example, the results on glycine and α -aminovaleric acid). Glycylphenylalanine is close to the other dipeptide listed, glycylglycine. Likewise, leucylglycylglycine is close to triglycine. The very close similarity between glycine ($^+\text{H}_3\text{N}\cdot\text{CH}_2\cdot\text{COO}^-$) and glycine betaine ($(\text{H}_3\text{C})_3\text{N}^+\cdot\text{CH}_2\cdot\text{COO}^-$), in which the three hydrogens attached to the nitrogen atom are replaced by three methyl groups, corresponds with the prediction of valence theory that in each case the positive charge will be effectively located on the nitrogen atom. The parallelism in the behavior of the three dichlorobenzenes and the three benzbetaines ($\text{COO}^- \cdot \text{C}_6\text{H}_4 \cdot \text{N}(\text{CH}_3)_3^+$) is significant, but it should be realized that in the former case the results stem from an *orientational* effect involving polar groups, and in the latter case from a distance effect involving *charged* groups.

At an early stage in the study of dipolar ions attention was called to a significant relation involving the dielectric increment which shows up in each of the two homologous series, that comprising the aliphatic amino acids and that comprising the glycine peptides. In each series the dielectric increment gives a good straight line when plotted against the length of the chain separating the charged groups in the molecule—i.e., in the case of the amino acids against the number of carbon atoms between the amino and carboxyl groups, and in the case of the peptides against the number of glycine units present. Both lines extrapolate back to zero at or near a point corresponding to zero length of the chain. Now we have seen that the dielectric increment is essentially proportional to the square of the dipole moment. At the same time, the moment itself is proportional to the distance between the charged groups. Consequently the observed relation implies that in each series the mean square value of the moment, or in other words the mean square distance between the charged groups, is proportional to the length of the intervening chain. This is essentially just what would be predicted on statistical grounds on the basis of free rotation. The problem of the mean square length of a molecular chain of identical atoms in which there is free rotation about the various valence bonds has been dealt with independently by Werner Kuhn and by Eyring (see for instance, Flory, 1953). By neglecting steric interference and any stabilizing interaction energy between different parts of the chain, Eyring obtains the result:

$$\overline{r^2} = l^2[n + 2(n-1)\cos\theta + 2(n-2)\cos^2\theta + \cdots + \cos^{n-1}\theta]$$

where, when $\overline{d^2}$ is the mean square distance between the centers of the terminal atoms of the chain, l is the distance between the centers of adjacent atoms of the chain, θ is the supplement of the valence angle, and n is one less than the total number of atoms in the chain. The series on the right may be summed to obtain the alternative equation:

$$\overline{d^2} = l^2 \left(\frac{n(1 + \cos \theta)}{1 - \cos \theta} \right) - 2 \cos \theta \frac{(1 - \cos^n \theta)}{(1 - \cos \theta)^2} \quad (115)$$

which for large values of n reduces to

$$\overline{d^2} = nl^2 \left(\frac{1 + \cos \theta}{1 - \cos \theta} \right) \quad (n \rightarrow \infty) \quad (116)$$

This is the same as the expression given by Kuhn on the basis of the same assumptions.

We have calculated the molar polarizabilities of the aliphatic amino acids and of the glycine peptides from the data in Table IV by equation (112) and compared them with the corresponding theoretical polarizabilities calculated on the assumption of unrestricted free rotation about the various valence bonds of each molecule, employing (115). Values for the α -, β -, γ -, and δ -amino acids were obtained by direct calculation on the following assumptions: the carbon-to-carbon distance is 1.54 Å; the carbon-to-nitrogen distance¹³ is 1.40 Å; the positive charge of the dipolar ion is located at the center of the nitrogen atom; the negative charge, as a result of resonance, is effectively located midway between the two oxygen atoms of the ionized carboxyl group and at a distance 2.14 Å from the α -carbon atom on the line joining that atom to the carboxyl carbon atom; the valence angle is the tetrahedral angle ($\cos \theta = -1/3$). Other values were calculated from equation (115), again taking the valence angle as equal to the tetrahedral angle, treating the nitrogen atom like a carbon atom, and considering the negative charge of the carboxyl group as if it were located at the carboxyl carbon atom. (The errors of these two assumptions tend to cancel one another and in any case have only an insignificant effect in the longer chain molecules.)

Both the data representing the glycine peptides and those representing the aliphatic amino acids show an almost perfect correlation between the measured polarizabilities and those reckoned from the mean square distance between the charged groups on the assumption of free rotation. In each case the points lie on a straight line which passes through the

¹³ The use of the more recent and more accurate value 1.47 for the C—N distance would hardly alter the results.

origin. In the case of the glycine peptides the slope of the line corresponds to

$$\frac{P_{\text{obs}} (\text{dielectric})}{P_{\text{calc}} (\text{free rotation})} = \frac{64}{40} = 1.6$$

If we assume that the value of P calculated on the assumption of free rotation corresponds to the square of the dipole moment μ_0 , and that obtained from the dielectric measurements corresponds to $\mathbf{u} \cdot \bar{\mathbf{u}}$, then it follows that

$$\mu_0 = 0.79 \sqrt{\mathbf{u} \cdot \bar{\mathbf{u}}}$$

In the case of the aliphatic amino acids the relation is found to be:

$$\frac{P_{\text{obs}}}{P_{\text{calc}}} = \frac{60}{40} = 1.5$$

which gives

$$\mu_0 = 0.81 \sqrt{\mathbf{u} \cdot \bar{\mathbf{u}}}$$

When we consider the simplifications employed in calculating the theoretical polarizabilities, either one of these two values may be considered to be in agreement with the conclusion reached in earlier parts of the discussion that for polar liquids generally we may expect μ_0 to be about 75% of $\sqrt{\mathbf{u} \cdot \bar{\mathbf{u}}}$. As a refinement the slight discrepancy between the values for the amino acids and those for the peptides might be ascribed to the fact that in calculating the theoretical polarizabilities of the peptides no account was taken of the contributions to the moments due to the peptide linkages. If we assumed that the effect of each such polar linkage was to contribute 3.5% to the net moment of the two adjacent amino acid residues, the two curves would be brought to coincide. This would correspond to a net vector contribution of about 0.8 Debye unit to the moment of a peptide by each peptide bond present in the peptide. This is considerably smaller than the inherent moment of the peptide bond, but the orientation of the latter is almost certainly not parallel to that of the molecule as a whole. An additional important factor is that the C—N bond of the (O=C—NH—) group in the peptides has a large amount of double-bond character, as we have already seen in Chapter 3. Hence there can be little or no free rotation around this bond, and this factor will certainly tend to increase the most probable end-to-end distance (\bar{d}^2) between the positive and negative charges at the ends of the peptide chains.

The results do not, of course, require that there is actual free rotation about all the bonds in the molecules. It may well be, as Mizushima has suggested on the basis of a study of infrared and Raman spectra, that there is a stabilization of the molecules in certain preferred configurations,

but the conclusion to which we are led is that the statistical result of this effect must be essentially equivalent to free rotation. Another closely related point is that there may be a tendency for the molecule, in the course of free rotation, to remain longer in configurations in which the oppositely charged groups are close to one another. If this were so, it would account for the fact that the ratio $\mu_0/\sqrt{\mathbf{u} \cdot \bar{\mathbf{u}}}$ as deduced in the preceding paragraph is slightly greater than what would be expected from data on pure liquids; but the effect appears to be small, and it would seem that we may be fairly confident in taking the intrinsic moment of a molecule as being close to 0.75 times that calculated as $\sqrt{\mathbf{u} \cdot \bar{\mathbf{u}}}$.¹⁴ Indeed such values are probably as close to the mark as is meaningful when we recall the fundamental nature of the concepts involved.

TABLE V
MOMENTS FROM DIELECTRIC INCREMENTS COMPARED WITH VALUES CALCULATED FROM FREE ROTATION

Substance	μ_0	$\sqrt{\mathbf{u} \cdot \bar{\mathbf{u}}}$	$\mu_0/\sqrt{\mathbf{u} \cdot \bar{\mathbf{u}}}$
Lysylglutamic acid	47.9	61.9	0.77
ξ -Aminopentadecylic acid betaine	39.5	54	0.73
π -Aminoheptadecylic acid betaine	42.1	51.3	0.82
Cystinyldiglycine	31.8	40.6	0.78
Diglycylcystine	31.8	40.6	0.78
Cystinyldidiglycine	40.9	53.7	0.76
ϵ, ϵ' -Diamino-di(α -thio- <i>n</i> -caproic acid)	35.1	39.8	0.88
ϵ, ϵ' -Diguanido-di(α -thio- <i>n</i> -caproic acid)	38.2	43.1	0.89

For references, see J. P. Greenstein and J. Wyman, *J. Am. Chem. Soc.* **58**, 463 (1936); J. P. Greenstein, J. Wyman, and E. J. Cohn, *ibid.* **57**, 637 (1935); J. P. Greenstein, F. W. Klemperer, and J. Wyman, *J. Biol. Chem.* **125**, 515 (1938); **129**, 681 (1939). For a general compilation of data, see E. J. Cohn and J. T. Edsall (1943), pp. 146-147.

In view of what has just been said, it is of interest to compare the moments of various other peptides, as calculated on the assumption of free rotation, with the observed values of $\mathbf{u} \cdot \bar{\mathbf{u}}$. This is done in Table V (see Wyman, 1939). Except for the last two entries in the table the results of the comparison accord with the results of the calculations on the simpler straight-chain compounds, according to which $\mu_0/(\mathbf{u} \cdot \bar{\mathbf{u}})^{1/2}$ is near 0.75.

¹⁴ Werner Kuhn has made a calculation of the effect of the electrostatic interaction of the charged groups on the mean square distance between them, assuming otherwise unhindered rotation. His results, in the light of the foregoing discussion, are almost certainly too small. Thus, according to them, the moment of tetraglycine should be only 24, as compared with the value $\mu_0 = 32$ which is three-fourths the value of $\sqrt{\mathbf{u} \cdot \bar{\mathbf{u}}}$ given in Table IV.

Finally, the extremely large polarizabilities of the proteins listed in Table IV deserve comparison with those of the amino acids. Even in the case of hemoglobin, the least polar, the polarizability, 5.8×10^6 cc, is about five times as great as the estimated value of 1.2×10^6 cc which it would have if it were a simple glycine peptide of the same molecular weight and fitted into the series of glycine peptides studied. The interpretation of the polarizabilities and moments of proteins, which, of course, unlike the glycine peptides, contain a large number of ionizable groups, will be discussed in Volume II.

Methods of Measuring Dielectric Constant

In order to connect the subject of this chapter more closely with experience it seems desirable to indicate in very general terms something of the nature of the operations by which the dielectric constant can actually be measured. These fall into two major categories, one involving determinations of the mechanical force between charged bodies, and the other involving the determination of electrical capacity. The second of these is the more important, but the first is perhaps conceptually simpler and we accordingly begin with it.

Conceptually, an ideal way to measure the dielectric constant of a medium would be to measure the force between two point charges immersed in the medium. From a knowledge of the magnitude and separation of the charges, the dielectric constant would follow at once from Coulomb's law. Actually such a procedure is not very practical. For one thing point charges are an abstraction, and for another thing it would be difficult to determine their absolute magnitude. To be sure we might employ small charges widely separated and compare the force between them first in a vacuum and second in the dielectric, but there would be the danger that the charges might change or leak away during the course of the operation, and in any case the measured force would be small.

An equivalent procedure which avoids the objections just mentioned is to measure the mechanical attraction between the plates of a condenser maintained at a given difference of potential by connecting it with the poles of a battery, first with the condenser empty and second with the condenser filled with (or immersed in) the medium in question. In this way, the forces may be made much bigger, and the problem of leakage of charge is eliminated even in a medium of appreciable conductivity. It is known from the principles of electrostatics that the energy density (energy per unit volume) stored at any point in an electric field is given by $E^2 D / 8\pi$, E being the electric intensity at the point. From this it follows at once that, if the plates of the condenser are maintained at a fixed difference of potential (which everywhere determines E) the force be-

tween them is proportional to the dielectric constant of the surrounding medium.¹⁵ For the condenser one may make use of the plates of a quadrant electrometer. Then the force between the plates, measured by the torsion on the suspension wire, is proportional to the dielectric constant of the medium surrounding the plates. This method of measuring the dielectric constant is known as the electrometer method. It is not quite so accurate as some other methods, but it is particularly useful in the case of conducting media.

A closely related method is the so-called ellipsoid method, which depends on the measurement of the torque acting on a conducting ellipsoid suspended in an electric field. This likewise has proved useful in the study of conducting media.

Measurements of the other category, involving determinations of capacity, depend on the fact that the capacity of a condenser is proportional to the dielectric constant of the medium with which it is filled or in which it is immersed. Nearly all measurements of this kind make use of alternating currents. To understand them, it is necessary to understand the concept of electrical impedance, which is a generalization of the concept of resistance that is fundamental to calculations involving alternating currents. A perfect condenser, of course, presents an infinite resistance to a direct voltage. Except for a momentary transient effect, involving the charging of the condenser to its equilibrium state, there is no flow of current into the condenser after application of the voltage. If, however, the voltage is made to alternate with sufficient speed, there will be a constant alternation of current flowing into and out of the condenser. There is a similar effect involving an inductance. For a steady voltage the inductance offers no resistance beyond the ordinary Ohm's law resistance of its wire to the flow of current. For an alternating current, however, there is at any moment a back voltage proportional to the instantaneous

¹⁵ This follows from the principle that the mechanical force in any direction is equal to the rate of change of the total energy of the system with a displacement in that direction. It is instructive to consider the case of a parallel plate condenser. By Gauss's law, the field between the plates of the condenser is uniform and given by $4\pi Q/A = DE$, where A is the area of the plates and Q is their charge. If l is the distance between the plates, the work of moving a charge dQ from one plate to the other in the direction opposite to E (i.e., so as to increase E) is $El dQ$. The total work of charging the condenser is therefore

$$\int_0^Q El dQ = \int_0^Q \frac{4\pi Q}{AD} l dQ = \frac{2\pi Q^2 l}{AD} = \frac{AlDE^2}{8\pi}$$

The volume between the plates of the condenser is Al , and the energy per unit volume is therefore $DE^2/8\pi$. The force of attraction between the plates is given by the rate of decrease of the total energy with l or $ADE^2/8\pi$ and is proportional to D .

rate of change of the current, the inductance acting to resist the alternation of the current. In each case, the "resistance" of the element, whether capacitance or inductance, to the alternating current is known as its impedance.

Let us reckon the impedance for each of these two circuit elements just mentioned, starting with the inductance. The voltage drop, V , produced by a current, I , flowing through an inductance, L , is given by

$$V = L \frac{dI}{dt} \quad (117)$$

where t denotes time. This voltage acts to oppose the current when the latter is increasing. Suppose that the current has the form of a simple harmonic oscillation:

$$I = I_0 \sin \omega t \quad (118)$$

where the period of alternation is given by

$$t = \frac{2\pi}{\omega} \quad (119)$$

Then it follows at once from (117) that

$$V = \omega L I_0 \cos \omega t = \omega L I_0 \sin \left(\omega t + \frac{\pi}{2} \right)$$

It will be seen that the effect of the inductance is to produce a voltage drop whose magnitude, $\omega L I$, is proportional to the current flowing, but which is out of phase with it, in fact leading it by the angle $\pi/2$.

If it were not for the complication involving the phase difference, we could simply regard the inductance as being equivalent to a resistance ωL . The phase shift, however, is an important feature of the effect. In order to include it, it is convenient to introduce complex quantities, which are of great utility in the discussion of all alternating current phenomena. Since e^{ix} is mathematically equivalent to $\cos x + i \sin x$, where $i = \sqrt{-1}$, we may write equation (118) in the form

$$I = I_0 e^{i\omega t}$$

with the understanding that I is to be identified with either the real or imaginary part of this complex expression. All the calculations may be carried through on this basis, provided in the end we make use of the real or imaginary part of the result. For the case in hand, we may write

$$V = i\omega L I = i\omega L I_0 e^{i\omega t} \quad (120)$$

If we identify I with the imaginary part of $I_0 e^{i\omega t}$ we see that this gives the equation arrived at above. (Similarly we might have identified I with the real part of $I_0 e^{i\omega t}$. Then the effect of the inductance would have been the same.) The expression $i\omega L$ which embodies the effect of the inductance both on the magnitude and on the phase of the voltage produced is known as its impedance, Z :

$$Z = i\omega L \quad (121)$$

It will be seen that the effect of multiplying a complex quantity e^{ix} by i is to produce an advance of phase of $\pi/2$, so that

$$V = i\omega L e^{i\omega t} = \omega L e^{i[\omega t + (\pi/2)]} \quad (122)$$

The expression for the impedance of a capacity may be derived in an analogous way. The voltage between the plates of a condenser, which opposes any further increase of charge, is given by $V = Q/C$, where C is the capacity and Q is the charge of the condenser. When differentiated with respect to time this gives

$$\frac{dV}{dt} = \frac{1}{C} \frac{dQ}{dt} = \frac{I}{C}$$

where I is the current flowing into the condenser in opposition to the voltage, V . If as before we write $I = I_0 e^{i\omega t}$, this gives

$$V = \int \frac{I_0 e^{i\omega t}}{C} = \frac{I}{i\omega C} = -\frac{iI}{\omega C}$$

Thus the impedance of a condenser is given by

$$Z = -\frac{i}{\omega C} \quad (123)$$

The condenser acts like a resistance equal to $1/\omega C$ and at the same time produces a *retardation* of phase $\pi/2$ in the voltage drop produced by a current flowing into it.

The quantities ωL and $1/\omega C$ of equations (121) and (123) are known as the impedance amplitudes, in the one case of the inductance, L , and in the other case of the capacitance, C . It will be seen that, as the frequency of an alternating current decreases to zero, one of these (ωL) goes to zero, as we know it should, and the other ($1/\omega C$) goes to infinity, also as we know it should. It is of interest to consider values of each of these quantities for several values of the frequency $\nu = \omega/2\pi$. In the case of the inductance the figures are for a coil having an inductance of 1 millihenry

such as is often used in radio sets, and in the case of the capacitance, for a 1-microfarad condenser such as is also commonly used.

IMPEDANCE (IN OHMS) OF A 1-MILLIHENRY INDUCTANCE AND A 1-MICROFARAD CONDENSER AS A FUNCTION OF FREQUENCY

ν	ωL	$1/\omega C$
10^3	2π	$2\pi \cdot 10^3$
10^6	$2\pi \cdot 10^3$	2π

It should be realized that impedances are composed according to the same principles as ordinary resistances. This arises from the fact that they enter into the equations for the flow of current in a circuit network in exactly the same way as simple resistances. Thus the total impedance of two impedances in parallel is given by

$$\frac{1}{Z} = \frac{1}{Z_1} + \frac{1}{Z_2} \quad (\text{parallel}) \quad (124)$$

and that of two impedances in series is given by

$$Z = Z_1 + Z_2 \quad (\text{series}) \quad (125)$$

For a circuit consisting of an inductance and capacitance in series we have from (125)

$$Z = i \left(\omega L - \frac{1}{\omega C} \right) \quad (126)$$

It will be seen that this goes to zero for $\omega L = 1/\omega C$ or

$$\omega^2 = \frac{1}{LC} \quad (127)$$

The frequency given by this equation is known as the resonant frequency of the circuit, and it is seen to be proportional to $1/\sqrt{LC}$. For an applied voltage of this frequency the circuit offers no resistance beyond the ordinary ohmic resistance of its wires and the current tends to become infinite, corresponding to the phenomenon of resonance. Resonance provides the basis for many of the most exact methods of measuring capacities and dielectric constants. Suppose, for example, we make use of a circuit in which the condenser containing the unknown dielectric is in parallel with a variable calibrated air condenser, and that the two are in series with an inductance. So long as the circuit remains in resonance at a fixed frequency, the sum total of the capacity of the two condensers must be constant. By introducing and removing the condenser containing

the unknown dielectric and adjusting the variable air condenser to maintain resonance it is possible, therefore, to measure the capacity of the former, and, consequently, to determine the dielectric constant of its contents directly.

Another way to proceed is to measure the resonant frequency of a rigid circuit embedded in the medium of unknown dielectric constant. The capacity of the whole circuit, and consequently the square root of the resonant frequency, is proportional to the dielectric constant. From the ratio of this frequency to the resonant frequency of the same circuit in a vacuum, the dielectric constant, therefore, is at once obtained. Determinations of frequency can be made with great accuracy.

It should be emphasized that no dielectric is a perfect insulator, as we pointed out earlier. All actual media possess some conductivity, and this will always be equivalent to a shunting resistance across the plates of a condenser in which they are present. It is important to see how this affects the impedance of the condenser. Let C be the capacity of the condenser and r the magnitude of the equivalent shunting resistance. Then, since the impedance of the resistance is, of course, simply r ,

$$\frac{1}{Z} = \frac{1}{r} - \frac{\omega C}{i}$$

Mathematically this is equivalent to

$$Z = \frac{r - ir^2\omega C}{1 + r^2\omega^2 C^2}$$

or

$$Z = \frac{1}{\omega C} \sqrt{\frac{r^2\omega^2 C^2}{1 + r^2\omega^2 C^2}} e^{-i \tan^{-1} r\omega C} \quad (128)$$

This shows that, when $r \rightarrow \infty$,

$$Z \rightarrow \frac{1}{\omega C} e^{i(-\pi/2)} = \frac{-i}{\omega C}$$

This corresponds to the case of a perfect condenser. When, on the other hand, $\omega C \rightarrow 0$, then $Z \rightarrow r$. Under intermediate conditions, when $r\omega C$ is finite, the effect of the shunting resistance, i.e., the conductivity of the dielectric, is to diminish the impedance amplitude of the condenser from the ideal value $1/\omega C$ which it would have in the absence of conductivity ($r \rightarrow \infty$) and to reduce the phase shift from its ideal value ($-\pi/2$). Moreover, if we write (128) in the form

$$Z = \frac{r}{\sqrt{1 + r^2\omega^2 C^2}} e^{i \tan^{-1} (-r\omega C)}$$

we see that as $r\omega C$ diminishes the impedance of the circuit approaches more and more nearly that of a pure resistance, r , and the effect of the capacity becomes a smaller and smaller part of the total.¹⁶ The result is that the resonance peak is progressively flattened and the accuracy of the determination of the resonance frequency falls off. To avoid these confusing effects, it is necessary to make the measurements at a sufficiently high frequency to keep $r\omega C$ large. This explains why so many dielectric constant measurements are made at high frequency.

Instead of basing the measurement of capacity on resonance, it is also possible to make a direct determination of the impedance of the condenser by introducing it into one arm of a Wheatstone bridge in which the signal is an alternating current of suitable frequency. Of course, in order to balance the bridge, it is necessary to introduce at least one other complex impedance into the bridge. We need not go into the matter in detail beyond pointing out that the calculations involved are formally the same as those used in the case of a simple direct-current bridge, the only difference being that the pure resistances are replaced by complex impedances. The bridge method gives both the resistance and the capacitance of the condenser. It has proved one of the most effective means of determining dielectric constants, but like the resonance method it is subject to difficulties when the medium has an appreciable conductivity. As the conductivity increases, the capacitance, which gives the dielectric constant, becomes a smaller and smaller part of the total quantity measured.

It has been shown that conductivity is a significant factor in the measurement of the dielectric constant. It is desirable, therefore, to relate the equivalent shunting resistance of a condenser with the conductivity of the dielectric. If we consider a parallel plate condenser filled with a medium of given dielectric constant and conductivity γ , the equivalent shunting resistance will be given by

$$r = \frac{l}{A\gamma} \quad (129)$$

where l is the distance between the plates and A their area. At the same time the capacity is given by

$$C = \frac{AD}{4\pi l} \quad (130)$$

The relation between r and γ is obtained by eliminating A/l between these two equations. It is

$$r = \frac{D}{4\pi\gamma C} \quad (131)$$

It can be shown (though we shall not do so here) that this relation is

¹⁶ For the exponential approaches unity, and so likewise does $\sqrt{1 + r^2\omega^2C^2}$

quite general. It should be realized, however, that in equation (130) C is given in electrostatic units, one of which is equal to 9×10^{11} practical units (farads). If C is expressed in farads, r in ohms, and γ in $\text{ohms}^{-1} \text{cm}^{-1}$, then equation (131) becomes

$$r = \frac{D}{4\pi\gamma C \times 9 \cdot 10^{11}} \quad (132)$$

This result will be useful later on in discussions involving anomalous absorption and power losses in liquids containing large molecules. It may be remarked in anticipation of such discussions that it follows from what has been said that any effect which leads to a phase difference other than $\pi/2$ between the current and the resulting voltage developed in a condenser will appear as a resistance effect. Thus, if large molecules lag behind the applied field in their orientation (state of polarization), the effect will show up in the measurements as a conductivity (resistance) effect.

SELECTED REFERENCES

- Böttcher, C. J. F. (1952). "Theory of Electric Polarisation," Elsevier Publishing Co., Houston, Texas.
- Cohn, E. J., and Edsall, J. T. (1943). "Proteins, Amino Acids and Peptides," Chapters 6 and 22, Reinhold Publishing Corp., New York.
- Debye, P. (1929). "Polar Molecules," Chemical Catalog Co., New York (recently reprinted by Dover Publications, New York).
- Flory, P. J. (1953). "Principles of Polymer Chemistry," Cornell University Press, Ithaca, New York. See especially Chapter X, Configurations of Polymer Chains.
- Fröhlich, H. (1949). "Theory of Dielectrics," Oxford University Press, London.
- Kirkwood, J. G. (1939). The dielectric polarization of polar liquids, *J. Chem. Phys.* **7**, 911.
- Onsager, L. (1936). Electric moments of molecules in liquids, *J. Am. Chem. Soc.* **58**, 1486.
- Partington, J. R. (1953, 1954). "An Advanced Treatise on Physical Chemistry," Vols. IV and V, Longmans Green and Co., London, New York, and Toronto. See especially Vol. V, pp. 287-541.
- Smyth, C. P. (1955). "Dielectric Behavior and Structure," McGraw-Hill Book Co., New York.
- Stuart, H. A. (ed.) (1952). "Die Physik der Hochpolymeren," Vol. I, "Die Struktur des Freien Moleküls," Chapter 5, Dielektrizitätskonstante, elektrisches Moment und Molekülstruktur, Springer-Verlag, Berlin.
- van Vleck, J. H. (1937). On the role of dipole-dipole coupling in dielectric media, *J. Chem. Phys.* **5**, 556.
- Wilson, J. N. (1939). The dielectric constants of polar liquids, *Chem. Revs.* **25**, 377.
- Wyman, J., Jr. (1936a). Polarization and dielectric constant of liquids, *J. Am. Chem. Soc.* **58**, 1482.
- Wyman, J., Jr. (1936b). The dielectric constant of solutions of dipolar ions, *Chem. Revs.* **19**, 213.
- Wyman, J., Jr. (1939). The problem of free rotation in complex dipolar ions in solution, *J. Phys. Chem.* **43**, 143.

Chapter 7

Conductivity of Electrolytes

Conductance and Resistance

It was pointed out in the preceding chapter that the practice often adopted in discussions of electrostatics, according to which all substances are classified as either insulators or conductors, is a gross oversimplification. All actual media have some degree of conductivity, though the difference between extreme cases, such as a gas on the one hand or a metal on the other, is enormous. The conductivity of any substance is measured, on the basis of Ohm's law, by the current which it passes when subjected to an electric field. Let us consider a rectangular block (parallelepiped) of the substance and suppose that a pair of opposite faces is subjected to a potential difference, ψ , by connection with the poles of a battery. Then by Ohm's law the current, I , which flows through the block is given by

$$I = \frac{\psi}{R} = \Gamma\psi \quad (1)$$

where R is the resistance of the block, and its reciprocal, Γ , is by definition the conductance of the block. It is obvious that the two quantities R and Γ depend not only on the intrinsic conductivity of the substance of which the block consists but also on its size and proportions. They may, however, be related very simply to other quantities not subject to this complication. It is found experimentally that Γ increases in proportion to the area, A , of the faces of the block, to which the potential difference is applied, and that it is inversely proportional to the distance, d , between them. Therefore, it follows that

$$R = \frac{d}{A} r \quad (2)$$

and

$$\Gamma = \frac{A}{d} \gamma \quad (3)$$

where r and γ are now quantities which depend only on the nature of the substance, either one being the reciprocal of the other. They are known respectively as the specific resistance and specific conductance of the

substance, and either one serves to characterize uniquely its power to conduct an electric current, under given conditions of temperature and pressure. They are to be identified with the resistance or conductivity of a unit cube of the substance. In what follows, we shall express the relations involved primarily in terms of conductance rather than of resistance.

TABLE I
DIMENSIONS OF CERTAIN ELECTRICAL QUANTITIES
(l denotes length, m mass, and t time; c denotes the
velocity of light = 3×10^{10} cm/sec)

Quantity	Electrostatic units	Electromagnetic units	Practical
Potential	$\psi: l^{1/2}m^{1/2}t^{-1}$	ψc	$\psi \cdot c \cdot 10^{-8}$
Current	$I: l^{3/2}m^{1/2}t^{-2}$	I/c	$I \cdot 10/c$
Resistance	$R: l^{-1}t$	Rc^2	$Rc^2 \times 10^{-9}$
Specific conductance	$\gamma: t^{-1}$	γ/c^2	$\gamma \times 10^9/c^2$

It will be seen from (2) and (3) that the dimensions of specific conductance are those of the reciprocal of a resistance multiplied by a length. It follows that if the resistance is expressed in ohms and the lengths in centimeters, then

$$\gamma = \frac{1}{\text{ohms cm}} \quad (4)$$

This is sometimes written as

$$\gamma = \text{mhos cm}^{-1} \quad (5)$$

the word mho being introduced for the unit of conductance corresponding to the ohm as the unit of resistance. In terms of fundamental units (mass, length, and time) the dimensions of γ will, of course, depend on the system of electrical units employed, since the dimensions of resistance are different for different systems. Since this is always a somewhat confusing matter, we introduce Table I, which shows the dimensions of the fundamental quantities ψ , I , R , and γ , according to each of the three principal systems of electrical units. For the most part we shall make use of practical units, and we shall express γ in terms of mhos cm^{-1} .

Electrolytes as Conductors

The concepts just developed are of general applicability. In this chapter from now on we shall be concerned with the conductivity of electrolytes. These represent a group of substances which occupy an intermediate position between insulators, on the one hand, and metals, on

the other. Thus, the specific conductivity of mercury at 0° is 10,630 mhos cm^{-1} , whereas that of a molar solution of potassium chloride, containing 71.14 grams per liter—a highly conducting electrolyte—is, at the same temperature, only 0.06517 mho cm^{-1} . The specific conductance of an organic liquid such as benzene or carbon tetrachloride is less, by many powers of 10, than that of the electrolyte. Whereas, in the case of the metals, lowering the temperature increases the conductance, in the case of electrolytes it decreases it. This, as well as the difference in the absolute values of the conductances, reflects the wholly different mechanisms involved in conductivity phenomena in the two cases. In the metals, it is the flow of free electrons which gives rise to the current; in electrolytes it is the migration of ions which produces the same result. Clearly a study of the conductivity of electrolytes cannot fail to shed much light on the behavior of ions in solution. The phenomena encountered will be a direct reflection of the degree of dissociation of the electrolyte molecules, of the interaction of the ions formed both with one another and with the solvent, and of forces acting on the ions as they move through the solution. In the case of large ions like proteins, a considerable part of what we know about the magnitude of the charges they bear, their isoelectric points, their heterogeneity, has come from a study of a phenomenon closely related to conductance, namely electrophoresis. The ideas presented in this chapter will provide the basis for a clearer understanding of these and other matters, as well as being of significance in their own right.

Faraday's Law

In order to study, indeed to realize, the phenomenon of electrolytic conduction, it is necessary to introduce a solution of the electrolyte as part of an electrical circuit containing a source of voltage. This involves two contacts between the solution and the metallic portions of the circuit. Each such contact necessitates the use of some kind of an electrode, for example, a bright platinum electrode or a silver chloride electrode. In the metallic parts of the circuit, as we have remarked, the current is due to the flow of electrons; in the solution, to the movement of ions. At the surface of an electrode the transfer of electricity is accomplished by an exchange of electrons which may involve either the production or discharge of an ion and constitutes a chemical reaction. For example, when positive current flows from a solution of silver nitrate into a silver electrode in contact with it, there is the production of a silver atom at the expense of a silver ion at the electrode surface. This represents the acceptance of an electron from the metallic silver electrode by an arriving silver ion. The silver atom produced by this reaction is deposited on the silver

electrode. If the current flows in the opposite direction, the reverse reaction occurs and a silver ion is produced in the solution at the expense of a silver atom lost by the silver electrode. Reactions of this sort associated with the flow of current are known as electrochemical reactions, and they always occur at electrode surfaces. In the body of the electrolyte there is simply a movement of ions, without chemical reaction. It was one of the great achievements of Faraday to show, in 1833, that the passage of a definite amount of electricity is always associated with a given electrochemical reaction at each electrode. In particular, for each equivalent (combining weight with hydrogen) which reacts there is always the transfer of 96,500 coulombs of charge. This is known as the faraday. It is clear from what has been said that it is the absolute amount of charge carried by one equivalent.

Let us now turn from a consideration of the reactions which occur at electrodes to the process of conduction within the body of the electrolyte, the subject with which we are primarily concerned.

Equivalent Conductivity

It will be seen from the discussion given in the opening section of this chapter that the specific conductance, γ , may be identified with the current which flows a unit distance through a unit area of an ideal plane surface in the medium normal to the electric field when the strength of the field is unity. (If, as usual, we express γ in mhos cm^{-1} , the unit area is 1 cm^2 and the unit field is 1 volt per centimeter.) As we have seen, in the case of an electrolyte the current is to be attributed to the movement of the ions. Part of it will be due to the migration of the positive ions to the cathode, and part to the migration of negative ions to the anode. In order to formulate this interpretation quantitatively, let us consider the case of a binary electrolyte which dissociates into ν_+ positive ions of valence z_+ and ν_- negative ions of valence z_- . Let us suppose that, under the influence of a unit electric field, these two kinds of ions move with the velocities u_+ and u_- . Let us suppose further that the concentration of the electrolyte, in moles per unit volume, is c , and that its degree of dissociation is given by the fraction α . It follows at once that

$$\gamma = Nc\alpha(\nu_+z_+\epsilon u_+ + \nu_-z_-\epsilon u_-) \quad (6)$$

where N is Avogadro's number and ϵ is the elementary charge. Now for reasons of electrical neutrality

$$\nu_+z_+ = \nu_-z_- \equiv n \quad (7)$$

The quantity n , which is, of course, unity for a uni-univalent electrolyte, is the number of equivalents associated with one mole of electrolyte,

and nc is, therefore, the number of equivalents of electrolyte in a unit volume of the solution.

The equivalent conductance is defined as

$$\Lambda = \frac{\gamma}{nc} \quad (8)$$

This means that it is the conductance per equivalent of electrolyte. It can be visualized as the conductance of a solution of concentration c , placed between two parallel electrodes 1 centimeter apart, the area of each being just sufficient so that there is 1 equivalent of electrolyte in the solution between the electrodes. In practice, of course, such an unwieldy system would never be used; we determine γ and c experimentally and calculate Λ from (8). If γ is expressed in mhos cm^{-1} , Λ has the dimensions of mhos cm^2 (gr equiv) $^{-1}$. If we introduce Λ into (6) we obtain the result

$$\Lambda = N\epsilon\alpha(u_+ + u_-) \quad (9)$$

or, taking account of the fact that $N\epsilon$ represents an amount of charge equal to the faraday,¹ which we denote by \mathbf{F} :

$$\Lambda = \alpha\mathbf{F}(u_+ + u_-) \quad (10)$$

This equation, stated in terms of the equivalent conductivity, Λ , is the embodiment of the physical interpretation of electrolytic conductivity as being due to the migration of ions, for the case of a binary electrolyte. It might, of course, be generalized to the case of more complicated electrolytes, but we shall not attempt to do so. It should be noted that the c involved in the definition (8) is concentration in moles per unit volume (cm^3). In terms of the concentration, C , in moles per liter the definition (8) becomes

$$\Lambda = \frac{1000\gamma}{Cn} \quad (11)$$

a form commonly given.

Ion Mobilities

The ion velocities u_+ and u_- which play such a fundamental role in the above analysis are known as the ion mobilities. It will be recalled that they are the velocities developed by the ions in the solution under the influence of a unit field (e.g., a field of 1 volt per centimeter). An ion moving in such a field is impelled by a force proportional to the intensity of the field. Its motion is resisted, owing to the viscosity (η) of the medium.

¹ The use of the symbol \mathbf{F} to denote the Faraday equivalent, here and in Chapter 8, should not be confused with its use in Chapter 6 to denote the internal field intensity.

If we picture the ion as resembling a macroscopic body moving through the liquid, then if the velocity of the ion is dx/dt , the resisting force is proportional to the velocity and is equal to $f(dx/dt)$. Here f is a frictional coefficient, proportional to the viscosity (η) of the medium, depending on the size and shape of the ion, including the molecules of solvent which are tightly bound to it and travel with it in its motion. For a sphere of radius r , f is equal by Stokes' law to $6\pi\eta r$. When the electric field is increased, the motion of the ion in the direction of the field is accelerated, and the viscous resistance to the motion increases as its velocity increases, until the resisting force just balances the force due to the field. This state of balance is attained in a minute fraction of a second after the field is applied; thereafter the net force on the ion is zero, and it moves through the medium with a constant velocity, equal to the field strength in volt cm^{-1} multiplied by the mobility.²

In the case of a strong electrolyte, for which $\alpha = 1$, if the sum of the ion mobilities, $u_+ + u_-$, were independent of concentration, the equivalent conductance, Λ , should be also independent of concentration. Actually this is never so, and in aqueous solutions there is always an increase of Λ with dilution toward a limiting value as illustrated by Fig. 1. This, of course, reflects changes in u_+ and u_- . Such changes may be attributed to interactions of the ions with one another and with the solvent which are dependent on concentration. This is a matter which will be discussed in a later section.

In the weak electrolytes, there is a far greater change of Λ with concentration than in the strong electrolytes. This is only what would be expected, since here the degree of dissociation, α , varies with dilution, as well as the interactions which affect the values of $u_+ + u_-$. In many earlier analyses of the change of Λ with concentration, such as those given by Wilhelm Ostwald, before the importance of interionic forces was appreciated, it was assumed that $u_+ + u_-$ remained constant at all concentrations. The entire variation of Λ with concentration was then ascribed to changes in the degree of dissociation, α , and was used, in some instances with considerable success, to calculate values of the dissociation constant of the electrolyte in question. (See Chapter 8, p. 421.) The variation of $u_+ + u_-$ with concentration found even in strong electrolytes shows,

² If the motion of the ion were very rapid, the resisting force could not be described simply as being proportional to the velocity; further terms in higher powers of the velocity would have to be introduced into the equations of motion. For the velocities attained in actual experiments on ions, however, these higher terms are completely negligible. It is because of this fact that it is possible to characterize an ion by a characteristic mobility, independent of the field strength, although dependent on the temperature and pressure and on the concentrations of the ions.

however, that in general such calculations should be accepted with caution and only as approximations.

It would be expected that at sufficiently high dilution the mobility of a given ion in a given solvent would have a characteristic value independent of the nature of the other ions present. Thus, we should expect that the mobility of the potassium ion in a very dilute solution of potassium chloride would be the same as its mobility in a very dilute solution of potassium sulfate, and the data bear this out. For example, consider

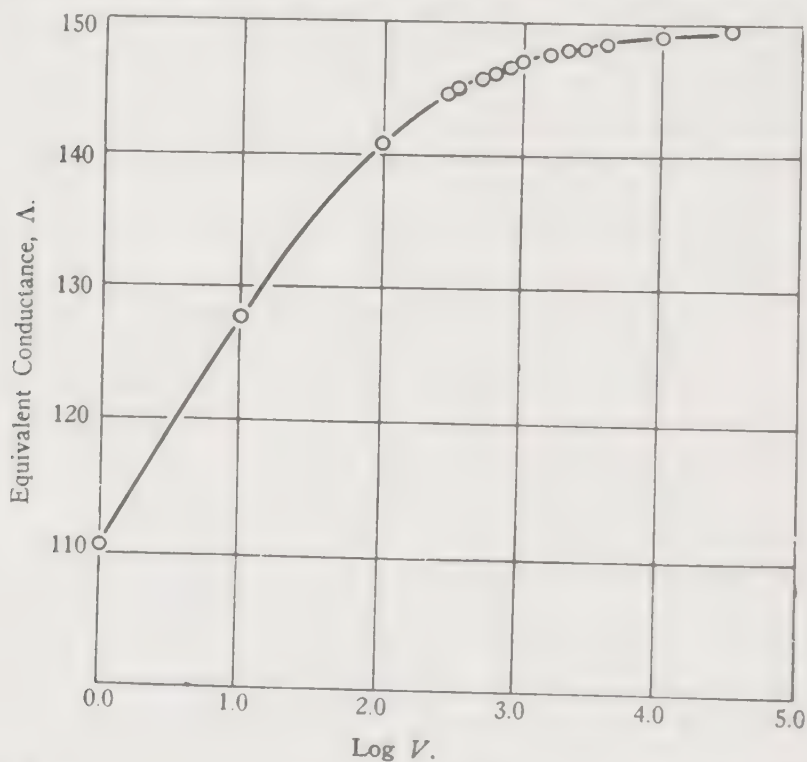


FIG. 1. Equivalent conductance of potassium chloride solutions as a function of V , the reciprocal of the molar concentration. (From D. A. MacInnes, "The Principles of Electrochemistry," 1939.)

the difference between the equivalent conductances of KNO_3 and NaNO_3 . This is found experimentally to be the same as the difference between the equivalent conductances of KCl and NaCl . This suggests that the mobilities of potassium and sodium ions are unaffected by the presence of the accompanying anions when the solutions are so dilute that interactions between the ions are negligible. On the basis of this, one might hope to be able to obtain limiting values of individual ion mobilities from data on the limiting values of conductances. This unfortunately is not so, for the number of mobilities involved always exceeds, at least by one, the number of measured conductances.

Ion Conductances

In order to obtain values of single ion mobilities, it is necessary to obtain further experimental information. This is provided by the study of what are known as transference numbers. Before we turn to this matter, however, there is one simple extension of our concepts which is worth introducing.

In the case of a strong electrolyte equation (10) may be written in the form

$$\Lambda = F(u_+ + u_-) \quad (12)$$

If as a matter of definition we introduce the quantities

$$\lambda_+ = Fu_+ \quad \text{and} \quad \lambda_- = Fu_- \quad (13)$$

this equation becomes

$$\Lambda = \lambda_+ + \lambda_- \quad (14)$$

Here λ_+ and λ_- are known as the ion conductances. If we introduce the subscript 0 to denote limiting values at infinite dilution, (14) becomes

$$\Lambda_0 = \lambda_{+0} + \lambda_{-0} \quad (15)$$

This represents Kohlrausch's law.

Transference Numbers

In the body of the electrolyte a certain fraction of the current which passes will be carried by each type of ion. This fraction is known as the transference number of the ion, and it is generally denoted by the symbol t . It can easily be shown, and is indeed almost self-evident, that for a binary electrolyte the transference number of an ion is the ratio of its mobility to the sum of its mobility and that of the other ion, e.g., $t_+ = u_+/(u_+ + u_-)$. The proof follows at once from equation (6), for the specific conductivity, γ , on the left of that equation may be identified with the current which flows a unit length through a unit area of a plane perpendicular to the direction of the applied field when the strength of the field is unity. Of this current an amount $(Nc\alpha\nu_+z_+eu_+)$ is carried by the positive ions, and an amount $(Nc\alpha\nu_-z_-eu_-)$ by the negative ions. Since by (7) $\nu_+z_+ = \nu_-z_-$, it follows that the fractions of the current carried by the positive and negative ions are given respectively by

$$t_+ = \frac{u_+}{u_+ + u_-} = \frac{\lambda_+}{\lambda_+ + \lambda_-} \quad (16)$$

and

$$t_- = \frac{u_-}{u_+ + u_-} = \frac{\lambda_-}{\lambda_+ + \lambda_-}$$

It is obvious that the two transference numbers t_+ and t_- are connected by the relation

$$t_+ + t_- = 1 \quad (17)$$

It will be seen that if we have any means of determining the transference numbers of the ions of a binary electrolyte, we can from the value of the equivalent conductivity determine the two ion mobilities. There are two principal methods for determining transference numbers; one is known as the Hittorf method and the other as the moving boundary method. We shall consider first the Hittorf method.

Hittorf Method of Determining Transference Numbers

Consider an electrolyte consisting of a solution of silver nitrate into which we dip two silver electrodes, one serving as the anode, one as the

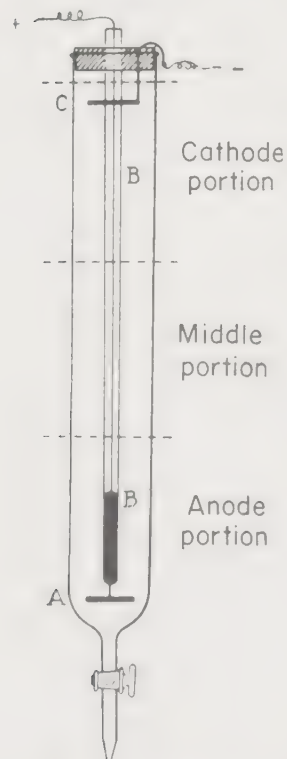


FIG. 2. A simple apparatus for the determination of transference numbers by the Hittorf method. (From MacInnes, 1939.)

cathode. When one equivalent (1 faraday) of charge flows through the solution, 1 gram atom of metallic silver is produced at the cathode and an equivalent amount of silver ion is produced at the anode, the reaction at the anode being the reverse of that at the cathode. Divide the solution into three ideal regions by two ideal fixed planes, which are shown in Fig. 2. We denote the region surrounding the anode by A , that surrounding the cathode by C , and that between them by B . The production of the silver atom at the cathode is associated with the loss of a silver ion by the solution contained in C . Similarly the production of a silver ion at the anode results in the gain of a silver ion by the solution contained in A . These changes, however, are not the only ones. The flow of one equivalent of charge through the solution involves the migration of t_{Ag^+} equivalent of silver ion from bottom to top through the middle portion and the migration of $t_{NO_3^-}$ equivalent of nitrate ion from right to left through the same region. As a result of this, there is no net gain or loss of either type of ion by the intermediate region B , where the composition of the solution, therefore, remains unchanged. In the region A , however, this migration of ions produces a loss of t_{Ag^+} equivalents of silver ion and a gain of $t_{NO_3^-}$ equivalents of nitrate ion.

If we take account also of the gain of one equivalent of silver ion due to the electrode process, we see that there is a net increase of silver ion in the region *A* which amounts to $(1 - t_{Ag}) = t_{NO_3}$ equivalent. This is exactly the same as the gain in nitrate ion, due to the migration above, as it surely has to be to provide for electrical neutrality. A similar analysis, applied to region *C*, shows that there is a net loss from that region of t_{NO_3} equivalents of both silver and nitrate ions. This again is as it has to be, since the silver nitrate which appears in *A* can only have come from *C*. We see, therefore, that if q equivalents of charge pass through the solution and Δ_{AgNO_3} represents the gain of silver nitrate in equivalents, in the region surrounding the anode, or the equal loss in the region surrounding the cathode, then the transport number of the nitrate ion is given by

$$\frac{\Delta_{AgNO_3}}{q}$$

It is a matter of simple analytical chemistry to determine Δ_{AgNO_3} ; q may be obtained by conventional electrical methods as the product of the current in amperes multiplied by the time during which the current flows.³

This example illustrates the principle of the Hittorf method of determining transport numbers, which was first introduced more than 100 years ago, and has provided one of the most important means of studying the conductance of electrolytes. It is significant that it long antedates the appearance of the Arrhenius theory of ionization in solution. There are, of course, experimental difficulties and complications in the application of the method, and many refinements have been introduced to increase its accuracy, but these need not concern us here. Instead of attempting to deal with them we shall pass on to the other principal method of determining transference numbers, namely the moving boundary method, which has now in large part replaced the older Hittorf method.

The Moving Boundary Method of Determining Transference Numbers

This method of determining transference numbers is based on the movement of the boundary between two electrolyte solutions under the influence of an electric field. The rate of movement of the boundary is a

³ In order to determine Δ_{AgNO_3} we measure the total amount of silver nitrate contained in a volume surrounding the electrode which is certainly large enough to include all regions where there is a change of concentration. We then compare this with the amount which would have been contained in the same volume if there had been no concentration change. It makes no difference just how large this volume is. It will be seen that the above expression always gives the transport number of the ion of opposite charge to that involved in the electrode reaction.

direct representation of the mobility of the ions and may be used to calculate their transference numbers. We have already explained above the intimate connection between ion mobilities and transference numbers.

Consider two solutions of two binary electrolytes, say potassium chloride and lithium chloride, having a common anion. Let us introduce these, without mixing, into a vertical tube, with the lighter solution on top to avoid convection. We suppose that by some suitable device it is possible to create a sharp initial boundary between them. (Such devices have been developed, and will be discussed later in connection with diffusion in Volume II.) Now let us pass a current through the tube in a direction such that the faster moving cation, in this case the potassium ion, is carried away from the initial position of the boundary. Then the boundary itself, which is made visible by the difference of refractive index

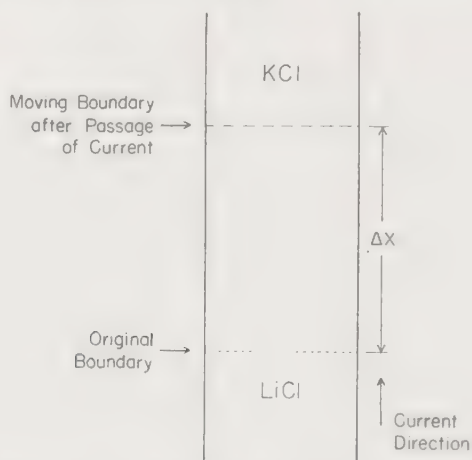


FIG. 3. The moving boundary principle, illustrated by a simple system. The direction of flow of the cations (positive current direction) is indicated by the arrow. The moving boundary has moved through the distance ΔX in time t , the volume swept out in this time being $A\Delta X$, where A is the cross-sectional area of the channel.

between the two solutions, will be found to move in the same direction and, if certain conditions are fulfilled, will remain sharp. (The nature of these conditions is further discussed below.) The boundary will continue to represent the interface between the two solutions. The rate at which it moves will, therefore, be the same as the actual rate of travel of either of the two different kinds of cation. Since these cations have different mobilities, the fact that they move at the same rate might seem surprising. The explanation is, of course, that, owing to the difference of conductivity, the strength of the field which establishes itself as a result of the applied voltage is different in the two solutions. The actual velocity of an ion is the product of its mobility and the field strength.

It is a very simple matter to develop an equation for calculating the

transference number of either cation on the basis of the velocity with which it moves. Let this velocity, which is, of course, nothing but the observed velocity of the moving boundary, be $\dot{x} = dx/dt$ cm sec⁻¹, where x is the coordinate measured along the axis of the tube. Let the cross section of the tube be A , and let the current which flows be I . In most modern experiments I is maintained at a constant value by some type of automatic control mechanism. Let the concentration of the cations in equivalents per cubic centimeter be c . Then the current carried by the cations is $FcA\dot{x}$. This may also be expressed as FcV/s , where $V = A\dot{x}$ is the volume swept out by the moving boundary in the time s seconds. But the ratio of this current to the total current, I , is by definition the transference number of the cation t_+ . Consequently, as our final equation, we have

$$t_+ = \frac{FcV}{Is} \quad (18)$$

The transference number of the corresponding anion may be obtained from this at once by the relation

$$t_+ + t_- = 1$$

In this example we have, in order to be specific, assumed that the two electrolytes have a common anion. We might, of course, equally well have assumed that the common ion was the cation.

At this point it behooves us to examine the basic condition which we have assumed to prevail in the above analysis, the condition that the boundary remains sharp and continues to separate the two solutions without any change in the composition of either. Since equation (18) is applicable to either cation (lithium or potassium) and since I and V are the same for both, it follows that

$$\frac{t_{\text{Li}}}{t_{\text{K}}} = \frac{c_{\text{LiCl}}}{c_{\text{KCl}}} \quad (19)$$

This represents a restriction in the relation between the concentrations of the two solutions. At the same time, it follows from the equal rate of travel of the lithium and potassium ions that

$$E_{(\text{in LiCl})}u_{\text{Li}} = E_{(\text{in KCl})}u_{\text{K}} \quad (20)$$

where the u 's denote mobilities as usual and the E 's represent the field strengths in the two solutions. But we also have from Ohm's law the condition that the E 's must be inversely proportional to the specific con-

ductances of the two solutions, γ_{LiCl} and γ_{KCl} . On the basis of (6) this gives the relation

$$\frac{E_{\text{LiCl}}}{E_{\text{KCl}}} = \frac{c_{\text{KCl}}(u_{\text{K}} + u_{\text{Cl}})}{c_{\text{LiCl}}(u_{\text{Li}} + u_{\text{Cl}})} \quad (21)$$

and this, in view of (16), is the same as

$$\frac{E_{\text{LiCl}}}{E_{\text{KCl}}} = \frac{c_{\text{KCl}}}{c_{\text{LiCl}}} \cdot \frac{t_{\text{Li}}}{t_{\text{K}}} \frac{u_{\text{K}}}{u_{\text{Li}}} \quad (22)$$

The three equations (19), (20), and (21) may be regarded as the mathematical expression of the conditions assumed in the analysis. It is easy to show that they are consistent.

A fundamental point which follows from this discussion is that any departure from the basic conditions assumed in our analysis tends to be self-correcting, so that the required physical situation tends to establish and maintain itself. Suppose that, owing to diffusion, one of the faster moving potassium ions wanders backward across the boundary into the lithium chloride solution, where the field is higher than on the other side. It is evident that as soon as it gets there it will tend to be expelled, being in its new position driven forward in the direction of the field with a velocity greater than that of the boundary. Similarly, any slow-moving lithium ion which strays into the potassium chloride solution will be quickly left behind and returned to the lithium chloride solution. Moreover, if the initial concentration of the lithium chloride solution is higher than the value required by (19), the field in that solution will be lower than that corresponding to (20), owing to higher conductivity, and there will be a tendency for lithium ions to lag behind until a steady state is realized. It is to these self-correction effects that a large part of the success of the moving boundary method must be ascribed.

In the above example we have said nothing about the chloride ions. It is evident that these will move in the opposite direction to that of the cations and, in each solution, with a velocity determined by the field strength. These velocities will, therefore, be different on the two sides of the boundary. There will, however, always be electrical neutrality in every element of volume of the solution.

It should be mentioned that in the above analysis we have neglected consideration of one possible source of error—namely, the volume changes which may occur at the electrodes. Such volume changes tend to displace the solution bodily and hence change the position of the boundary, quite apart from its motion due to the transport of the ions. This effect is not difficult to correct for, but we shall not attempt here to go into the matter. (See, for instance, MacInnes, 1939, Chapter 4.)

The situation discussed here corresponds to the simplest type of moving boundary system, in which there are only three ions of different mobilities, with one moving boundary and one approximately stationary boundary, at or near the original position from which the moving boundary started. The presence of this stationary boundary is shown by the existence of a gradient of concentration of the salt which remains at the level of the initial boundary—in this case lithium chloride. We have not hitherto referred to this boundary, since its presence need not be considered in the determination of transference numbers.

In the more general case of a system containing p anions and q cations, the theoretical treatment developed independently by V. P. Dole and by H. Svensson shows that there should be $p - 1$ boundaries with negative velocities (i.e. moving against the positive current), $q - 1$ boundaries with positive velocities, and one stationary boundary. A good brief presentation of this more general theory, with references, is given by Alberty (1953); and we shall discuss the problem in detail in connection with the treatment of electrophoresis in Volume II.

In Table II are listed the limiting equivalent conductances, at infinite dilution, for various important ions at 0° , 25° , and 45° . These were determined from conductance measurements and measurements of transference numbers. The equivalent conductance of any salt or acid, composed of any combination of these ions, is given, at infinite dilution, by equation (15). The mobilities of the ions are related to the ion conductances by (10), and the transference numbers at infinite dilution by (16). All these quantities are thus readily derived from the data of Table II.

We may offer some comments on Table II, expressing the data in terms of the ionic mobilities rather than the conductances. The limiting mobilities, at infinite dilution, of most small ions, in water at 25° , are of the order of $5 \times 10^{-4} \text{ cm}^2 \text{ volt}^{-1} \text{ sec}^{-1}$. That is, such an ion would move at the rate of $5 \times 10^{-4} \text{ cm sec}^{-1}$, when the applied field is 1 volt cm^{-1} . For Li^+ at 25° , for instance, u is 4.03×10^{-4} ; for Na^+ , 5.20×10^{-4} ; for K^+ , 7.62×10^{-4} ; and for NH_4^+ , almost the same as for K^+ . For some important simple anions, values at 25° are: Cl^- , -7.9×10^{-4} ; acetate, -4.24×10^{-4} ; and bicarbonate, -4.61×10^{-4} . Change of temperature greatly affects the mobility, chiefly by its effect on the viscosity of the solvent. Thus, the viscosity of water is 0.894 centipoise at 25° and is almost exactly twice as great (1.792) at 0° . Likewise, the mobility of the Li^+ ion is almost exactly twice as great at 25° as at 0° . The change is somewhat less for many ions; thus, the mobility of K^+ only increases by about 85% between 0° and 25° , instead of doubling as for Li^+ .

The mobilities of H^+ and OH^- are of a higher order of magnitude than those of any other ions known. That of H^+ is 36.3×10^{-4} ; that of OH^- ,

20.6×10^{-4} at 25° . Here the mechanism of transport in the solution is different from that of most ions; it involves the jumping of protons from one water molecule to another—a sort of chain mechanism of proton transfer through the solvent, which we have already briefly discussed in Chapter 2. Such a process naturally encounters much less resistance than the motion of a larger ion as an entire unit through a viscous medium.

TABLE II
IONIC EQUIVALENT CONDUCTANCES (λ_{0+} AND λ_{0-}) FOR SOME
IMPORTANT IONS AT 0° , 25° , AND 45°

Ion	Temperature, $^\circ\text{C}$		
	0°	25°	45°
H^+	225	349.8	441.4
OH^-	105	198.6	
Li^+	19.4	38.7	58.0
Na^+	26.5	50.1	73.7
K^+	40.7	73.50	103.5
NH_4^+	40.2	73.55	
CH_3NH_3^+		58.7	
$(\text{CH}_3)_3\text{NH}^+$		47.2	
$(\text{CH}_3)_4\text{N}^+$	24.1	44.9	
Mg^{++}	28.9	53.0	
Ca^{++}	31.2	59.5	88.2
Sr^{++}	31	59.5	
Cl^-	41.0	76.35	108.9
Br^-	42.6	78.1	110.7
I^-	41.4	76.8	108.6
NO_3^-	40.0	71.46	
CH_3COO^-	20.1	40.9	
Oxalate $^{--}$		74.1	

The data in this table are taken from R. A. Robinson and R. H. Stokes, "Electrolyte Solutions," Academic Press, New York, and Butterworths Publications, London, 1955.

The viscosity of water in centipoise is: at 0° , 1.792; at 25° , 0.894; and at 45° , 0.599.

Very large ions, such as proteins, have larger frictional coefficients and correspondingly lower mobilities than smaller ions. Thus, the mobility of human carbonyl-hemoglobin (molecular weight 66,700) is near $+2.5 \times 10^{-6} \text{ cm}^2 \text{ volt}^{-1} \text{ sec}^{-1}$ at pH 5.8, and -2×10^{-5} at pH 8, in phosphate buffer at 0.1 ionic strength at 4° . The isoelectric point, of zero mobility,

is near pH 7, and the net charge on the molecule (Z) changes by approximately 10 proton units per pH unit in this range. Any precise comparison with simple ions is difficult, but it is apparent that the ratio of mobility to net charge is much smaller for macromolecules than for the simple ions we have been considering.

Theoretical Considerations Regarding Equivalent Conductance; Effects of Interionic Forces

It was pointed out earlier in this chapter that even in the case of strong electrolytes where the degree of dissociation may be taken as unity there is a very significant change of equivalent conductance, Λ , with concentration. In aqueous solutions Λ increases with dilution to an upper limiting value, and in the range of fairly high dilutions is found to be linear in the square root of the ionic strength. An explanation of this behavior, based on the concepts of the Debye-Hückel theory of ionic interactions, has resulted from the theoretical investigations of Debye and Hückel, of Falkenhagen, and of Onsager. The subject is a difficult one and lies outside the scope of this book. We shall attempt no more, therefore, than to give a most general indication of the nature of the physical ideas involved and to present governing equations arrived at for the case of a binary electrolyte.

As in Chapter 5, we may choose any ion as the central ion in our system of coordinates, and consider the ion atmosphere around this ion taken as a fixed origin. We consider the mobility of the central ion, and the effects produced by the ion atmosphere on it. These effects are of two sorts. The first is known as the electrophoretic effect and results from the movement of the ion atmosphere as a whole under the influence of the applied electric field. It was first recognized and investigated by Debye and Hückel. The second is known as the time of relaxation effect. It has its origin in the disturbance of the symmetrical relation between the central ion and its atmosphere resulting from the movement of the central ion. It has been exhaustively studied by Onsager, whose limiting law for the variation of equivalent conductance with ionic strength takes account of both effects.

Let us consider first the electrophoretic effect. The total net charge of an ion atmosphere is equal and opposite to that of the central ion. If, therefore, under the influence of the applied field the central ion tends to move to the right, the ion atmosphere tends to move to the left. As it does so, it tends to carry the surrounding solvent with it. The central ion, therefore, finds itself in a moving stream of solution which bears it in a direction opposite to that in which it travels under the influence of the field. The result of this is to diminish its net velocity. As the ionic strength

becomes smaller, the ionic atmosphere, of course, fades out and recedes from the central ion. Hence, the effect becomes smaller and smaller as the ionic strength decreases until, in an infinitely dilute solution in the pure solvent, the final limiting value of the conductance is achieved. It is striking that an essentially correct qualitative formulation of this simple picture can be obtained by treating the ion atmosphere as if it were a uniformly charged shell, moving, with all the solvent which it contains, as a rigid body through the solution. In doing this it is only necessary to take the radius of the shell as equal to $1/\kappa$, where κ is the parameter of the Debye-Hückel theory (Chapter 5, equation 65), which is a measure of the size of the ion atmosphere, and to assume the charge of the shell to be the same as the net charge of the ion atmosphere. The result thus shows that the velocity, \dot{x}' , of the shell, and consequently of the moving fluid in which the central ion finds itself, is proportional to κ , that is, to the square root of the ionic strength:

$$\dot{x}' = \frac{E\epsilon\kappa}{6\pi\eta}$$

Here η is the viscosity of the medium, and ϵ is the charge of the central ion, equal, of course, to the charge of the ion atmosphere. On the other hand, the velocity of the central ion through the surrounding fluid is:

$$\dot{x} = \frac{E\epsilon}{6\pi\eta r}$$

where r is the radius of the ion. The ratio $\dot{x}'/\dot{x} = \kappa r$ is thus proportional to the square root of the ionic strength.

The basic idea of the time of relaxation effect is that there will be a disturbance of the symmetry of the ion atmosphere about the central ion as the latter moves under the influence of the applied field. This is related to the relaxation time of the atmosphere—a quantity which, roughly speaking, is a measure of the time required for the atmosphere to adjust itself to altered conditions—in this case to a change of the electric field resulting from the motion of the central ion. It can be predicted that the relaxation time, τ , of the ion atmosphere will be of the order of

$$\tau = \frac{f}{\kappa^2 k T}$$

where f is the frictional force constant of a single ion, and κ is again the parameter of the Debye-Hückel theory. Owing to its finite relaxation time, the ion atmosphere will lag behind the moving central ion by a distance of the order of $\tau\dot{x}$, \dot{x} being the velocity of the ion, and will be

deformed from its original spherically symmetrical shape. The result of this is that it will exert a backward pull on the central ion whose magnitude, in a medium of dielectric constant D , may be estimated as

$$\Delta\mathfrak{F} = \frac{\epsilon^2 \kappa^3}{D} \frac{\tau \mathfrak{F}}{f} = \frac{\epsilon^2 \kappa \mathfrak{F}}{DkT} \quad (23)$$

where f is again the friction force constant for an ion, and \mathfrak{F} is the force exerted on the central ion by the applied field. It will be seen that $\Delta\mathfrak{F}$, the backward pull exerted by the ion atmosphere, is proportional to κ and consequently to the square root of the ionic strength.

A rigorous calculation of both effects, the electrophoretic effect and the time of relaxation effect, leads to the limiting law for the equivalent conductance of electrolytes in fields of low intensity and low frequency, due to Onsager. For a binary electrolyte this law is given by the equation

$$\Lambda = \Lambda_0 - S \sqrt{\omega} \quad (24)$$

where ω is the ionic strength expressed in terms of concentrations in moles per liter.⁴ The proportionality constant, S , of this equation is of a complicated form, being given by

$$S = \frac{2.801 \times 10^6}{(DT)^{3/2}} \left(\frac{q}{1 + \sqrt{q}} \right) |z_1 z_2| \Lambda_0 + \frac{41.25(|z_1| + |z_2|)}{\eta(DT)^{1/2}} \quad (25)$$

where

$$q = \frac{|z_1 z_2|}{(|z_1| + |z_2|)} \frac{\lambda_{10} + \lambda_{20}}{(|z_2| \lambda_{10} + |z_1| \lambda_{20})}$$

Here the quantities z denote valencies as usual, and λ_{10} and λ_{20} are the limiting conductances of the two kinds of ion ($\Lambda_0 = \lambda_{10} + \lambda_{20}$). For a uni-univalent electrolyte like KCl, the expression for S degenerates into the much simpler form:

$$S = \frac{8.206 \times 10^5}{(DT)^{3/2}} \Lambda_0 + \frac{82.50}{\eta(DT)^{1/2}} \quad (26)$$

The approximations made in deriving Onsager's limiting law, expressed by (24) and (25), render it valid only at high dilutions. These approximations are least serious for uniunivalent electrolytes; and for these (24) and (26) describe the equivalent conductance measurements with high accuracy at concentrations below about 0.001 M . The experi-

⁴ $\omega = \frac{\sum C_i z_i^2}{2}$

mental data can often be fitted well, up to considerably higher concentrations, by a simple modification of (24):

$$\Lambda = \Lambda_0 - S \sqrt{\omega}/(1 + \kappa a) \quad (27)$$

Here κ is again the Debye-Hückel term, which is proportional to the square root of the ionic strength; and a is the mean collision diameter of the ions (Chapter 5, Fig. 11) which may be treated as an empirical parameter in fitting experimental conductivity data to equation (27). Theoretical justification for the introduction of the term $(1 + \kappa a)^{-1}$ into (27) has been provided by recent work of H. Falkenhagen and others. The situation has been well discussed by Robinson and Stokes (1955, Chapter 7), who give many references to the more recent work.

As we have indicated, the Onsager limiting law applies to conductances measured at low field strength and low frequencies. In very intense fields, of the order of several hundred thousand volts per centimeter, the conductances of electrolyte solutions are no longer constant but increase with the applied field. In such cases Ohm's law obviously breaks down. The ions move so rapidly in such fields that they at least partly escape from the retarding influence of the ion atmosphere. In addition, weak electrolytes show a markedly increased degree of ionization, as if the field were pulling the ions apart. These results are known collectively as the *Wien effect*. Special techniques are required in its study, for the applied fields can be employed only for very brief periods—of the order of 10^{-6} second—or else the heating effects produced on the liquid would become intolerable. The use of the Wien effect in determining the true ionization constant of carbonic acid is briefly discussed in Chapter 10.

In fields of ordinary intensity, but very high frequency, relaxation effects are observed, and the conductance becomes complex. These matters, however, lie outside the scope of this chapter.

Measurement of Conductivity

We shall not attempt to discuss this subject, except to point out that in order to avoid the cumulative disturbing effect of reactions at electrodes (so-called polarization effects), nearly all measurements of the conductivity of electrolytes are made with alternating currents. For the most part they involve the use of a bridge. A very rough indication of the concepts involved in such measurements was given at the end of the last chapter in the section dealing with measurement of dielectric constants.

REFERENCES

- Alberty, R. A. (1953) "Electrochemical Properties of the Proteins and Amino Acids" in "The Proteins" (H. Neurath and K. Bailey, eds.), Vol. IA, Chapter 7; see especially pp. 523-529.

- Harned, H. S., and Owen, B. B. (1950). "The Physical Chemistry of Electrolytic Solutions," 2nd ed., Reinhold Publishing Corp., New York. More advanced in character and more difficult than MacInnes; many valuable tables of data.
- MacInnes, D. A. (1939). "The Principles of Electrochemistry," Reinhold Publishing Corp., New York. Note especially Chapters 2, 3, 4, 18, 19, and 20. A lucid and excellent exposition.
- Robinson, R. A., and Stokes, R. H. (1955). "Electrolyte Solutions," Academic Press, New York, and Butterworths Scientific Publications, London. Contains many new references, including some very important ones not given in the other two books given directly above. Valuable discussion of many points in theory of electrolytes.

Chapter 8

Acid-Base Equilibria

The Nature of Acids and Bases

The concepts of acidity and basicity, like those of oxidation and reduction, have passed through a long process of evolution. The simplest criterion of acidity in a moderately acid solution is that it is sour, like vinegar or lemon juice. Alkalies, like the caustic potash obtained from wood ashes, are bitter; they react with acids with the evolution of a large amount of heat. If the two reactants are mixed in the right proportions, the product of the reaction is neither sour nor bitter, but salty in taste.

From these simple and primitive experiences, more general conceptions of acids and bases have gradually evolved. During the nineteenth century, it became apparent that all substances commonly called acids contain hydrogen. As knowledge of the existence of ions in solution developed, and with deeper understanding of the nature of atoms and molecules, it gradually became apparent that this hydrogen must be present in such a form that it is readily removable as a hydrogen ion or proton. Hydrogen attached to carbon is rarely acidic to any appreciable degree in aqueous solution, although occasionally it is so, as in HCN . Protons attached to nitrogen, oxygen, sulfur, halogens, and certain other atoms, however, are very commonly capable of being readily removed, and the structures to which they are attached thus function as acids. If an acid be denoted as HR , then, following J. N. Brönsted, we may denote the structure, R , which remains after a hydrogen ion has been removed, as the base which is conjugate to the acid HR . Since the hydrogen ion is positively charged, it is clear that either HR or R , or both, must be electrically charged, and further that the net charge on R must be more negative (or less positive) than that on HR by one proton unit (4.802×10^{-10} esu). If the net charge of HR is Z , then the charge of R is $Z - 1$. Either an acid or a base may be a cation, an anion, or an uncharged molecule. Some examples of acids and their conjugate bases are listed in Table I. The pK_A values, which are an index of the relative strength of these acids in water, are also listed (see the later discussion); the higher the value of pK_A , the weaker the acid.

It is apparent from Table I that many substances can function as both acids and bases—that is, they can be both proton donors and proton ac-

TABLE I
 SOME ACIDS AND THEIR CONJUGATE BASES

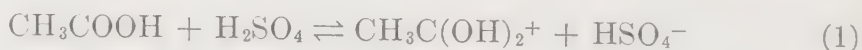
Acid	Conjugate base	$pK_A \equiv -\log K_A$ in water at 25°
H_3PO_4	$H_2PO_4^-$	2
$H_2PO_4^-$	HPO_4^{--}	7
HPO_4^{--}	PO_4^{---}	12
H_3O^+	H_2O	—
H_2O	OH^-	15.7
CH_3COOH	CH_3COO^-	4.75
NH_4^+	NH_3	9.3
$CH_3NH_3^+$	CH_3NH_2	10.7
$HOOC \cdot CH_2 \cdot COOH$	$HOOC \cdot CH_2 \cdot COO^-$	2.8
$HOOC \cdot CH_2 \cdot COO^-$	$^-OOC \cdot CH_2 \cdot COO^-$	5.7
$^+H_3N \cdot (CH_2)_2 \cdot NH_3^+$	$^+H_3N \cdot (CH_2)_2 \cdot NH_2$	7.0
$^+H_3N \cdot (CH_2)_2 \cdot NH_2$	$H_2N \cdot (CH_2)_2 \cdot NH_2$	10.0
$^+H_3N \cdot CH_2 \cdot COOH$	$^+H_3N \cdot CH_2 \cdot COO^-$	2.3
$^+H_3N \cdot CH_2 \cdot COO^-$	$H_2N \cdot CH_2 \cdot COO^-$	9.7
$\begin{array}{c} HC=CH \\ \quad \\ ^+HN \quad NH \\ \diagdown \quad / \\ C \\ \\ H \end{array}$ Imidazolium ion	$\begin{array}{c} HC=CH \\ \quad \\ N \quad NH \\ \diagdown \quad / \\ C \\ \\ H \end{array}$ Imidazole	7.0
$\begin{array}{c} H_2^+N \quad NH_2 \\ \diagdown \quad / \\ C \\ \\ NH_2 \end{array}$ Guanidinium ion	$\begin{array}{c} HN \quad NH_2 \\ \diagdown \quad / \\ C \\ \\ NH_2 \end{array}$ Guanidine	14 (approx.)
$\begin{array}{c} OH \\ \\ \text{C}_6\text{H}_5 \end{array}$ Phenol	$\begin{array}{c} O^- \\ \\ \text{C}_6\text{H}_5 \end{array}$ Phenolate ion	10.0

The pK_A values shown are an index of the relative strength of the acids shown, in water at 25°; the smaller the pK_A value, the stronger the acid.

Some of the structures shown are incomplete, in that they show only one out of several resonance components making up the actual structure. This is notably true of the guanidinium and the imidazolium ions, and of their conjugate bases. The guanidinium ion, for instance, is a planar symmetrical structure in which the three C—N bonds are all equivalent, with angles of 120° between them. This equivalence is destroyed when a proton is lost, and the guanidinium ion is converted to guanidine. The accompanying decrease of resonance energy in the basic form tends to stabilize the conjugate acid, and partly explains why the guanidinium ion is such a weak acid. Similarly the imidazolium ion has higher symmetry, and higher resonance energy, than basic imidazole. For further discussion see, for instance, Pauling (1940, Chapter VI); Cohn and Edsall (1943, Chapter 5).

ceptors. The most important of all these substances is water, which, acting as a base, gives the conjugate acid H_3O^+ ; as an acid, it gives the conjugate base OH^- . Other notable examples are the primary and secondary phosphate ions, H_2PO_4^- and HPO_4^{2-} ; the $\text{HOOC}\cdot\text{CH}_2\cdot\text{COO}^-$ ion, derived from malonic acid by loss of a proton; the $^+\text{H}_3\text{N}\cdot(\text{CH}_2)\text{NH}_2$ ion, derived from ethylenediamine by addition of a proton; and the dipolar ion glycine, $^+\text{H}_3\text{N}\cdot\text{CH}_2\text{COO}^-$. The class of macromolecules known as polyelectrolytes consists of substances having very large numbers of acid or basic groups per molecule. The proteins and nucleic acids are in this class, and we shall consider such molecules in more detail in Chapter 9.

In terms of modern valence theory, one characteristic of all bases may be noted. A base must possess at least one pair of *unshared* electrons; the base thus acts as a proton acceptor and an electron donor. All bases are electron donors; but the converse is not true. Many substances containing unshared electron pairs show no appreciable tendency to attach a proton and hence cannot be considered as bases. Substances like urea, however, which in water are exceedingly weak bases, often show well-marked basic properties in media like sulfuric acid or glacial acetic acid. Even acetic acid can act as a base in sulfuric acid, according to the equation

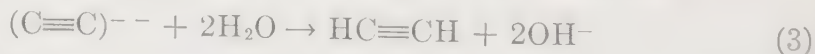


In liquid ammonia, NH_3 can also act as both acid and base:



This is exactly analogous to equation (12) for water, given below.

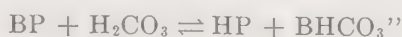
In the widest sense, indeed, every proton in any molecule is potentially removable as such, and the molecule is therefore acidic. The acid properties of acetylene, for instance, are well known; its calcium salt, calcium carbide, in the solid state has been shown to have an ionic structure $(\text{Ca}^{++})(\text{C}\equiv\text{C})^{--}$. In the presence of water the $(\text{C}\equiv\text{C})^{--}$ ion acts as an extremely strong base, forming acetylene and hydroxyl ion:



Ethylene is a still weaker acid than acetylene, and the saturated hydrocarbons are weaker yet. All the hydrocarbons, indeed, are such weak acids that their potential acidity can be completely ignored in the study of biochemical systems. The study of acidity in nonaqueous systems, however, has greatly broadened our conceptions of the nature and scope of acid-base equilibria. In what follows we shall be concerned principally with aqueous solutions, with some excursions into systems of lower dielectric constant, such as alcohols, or mixtures of water with such solvents as alcohol, acetone, or dioxane. The reader should remember, however,

that the domain of the subject of acid-base equilibria is much broader than this.

There is another use of the term "base," entirely different from that employed here, but widespread in the literature of clinical chemistry, and also employed by L. J. Henderson in his classic book on blood. Thus Henderson ("Blood," p. 60) writes: "Physiologists have long believed that the proteins of blood combine with base, that the quantity of base thus bound is variable, and that this combination takes part in the transport of carbonic acid, through changes in the equilibrium of the reaction,



In the current language of physical and organic chemists, we may translate Henderson's term "base" as meaning "cations," especially inorganic cations such as sodium, potassium, calcium, and magnesium. (These are sometimes referred to also as "fixed bases" because they are not volatile on heating in an oven up to several hundred degrees centigrade, whereas salts containing such cations as ammonium ion are volatile.) In the language of the Brönsted-Lowry theory, which we employ in this book, the protein P in the above equation is a base, since it combines with a proton to form the acid HP; also, of course, the bicarbonate ion (HCO_3^-) is a base, conjugate to H_2CO_3 .

The use of the word "base" for cations is a survival from the terminology of the late eighteenth and early nineteenth centuries when, in the study of the oxides of the metals, the metal was referred to as the "base" or principal important component of the oxide. At that time, of course, the distinction between metals and metallic ions was unrecognized, since the ionic theory did not arise until 1887 in the work of Arrhenius, and was not fully assimilated into the thinking of chemists until long after that. Indeed, during the period when Henderson's fundamental studies on buffer action and his work on blood were carried out, salts such as sodium chloride or bicarbonate were considered as being only partially dissociated in water, so that it was natural to write BHCO_3 (as in the quotation from Henderson above) for the bicarbonate of an ion such as sodium or potassium. It has since been shown that the activity coefficients of such salts in dilute aqueous solution are adequately explained by considering them as consisting of ions, which exert electrostatic forces on one another—see the discussion of the Debye-Hückel theory in Chapter 5. Hence today we write the bicarbonate of an univalent cation as B^+HCO_3^- , and regard the cation as an independent kinetic entity in solution, perhaps surrounded by a shell of closely attached water molecules, but certainly not as being bound to the bicarbonate ion—merely drawn toward it by electrostatic attraction. The cation as such is not involved in acid-base equilibria.¹

¹ Multivalent cations, surrounded by a firmly attached shell of water molecules, may sometimes act as acids by causing the ejection of a proton from one of these bound water molecules, as for instance with ferric iron in water. There is good reason to believe that the Fe^{+++} ion actually exists in water as $\text{Fe}(\text{H}_2\text{O})_6^{+++}$, the oxygens of the six water molecules being arranged at the corners of a regular octahedron around the Fe^{+++} ion. This may react as an acid according to the equation



Zinc, copper, chromium, manganese, and other cations undergo similar reactions, the resulting hydroxides being often quite insoluble in water, so that the acid pK values involved may be difficult to determine.

The inquiring student of biochemistry should be familiar with this traditional use of the term "base" to denote cations, so that he may properly understand the important writings of investigators as Henderson, D. D. Van Slyke, J. L. Gamble, J. P. Peters, and others, and can correlate their terminology with that currently used by physical and organic chemists.

Still another use of the term "base" has been widely employed in the past. For ammonia, the amines, and many other organic bases, the results of many acid-base equilibrium studies have been expressed in terms of a "basic dissociation constant," K_B . For such a base in water—for instance for a primary amine, RNH_2 —we may write the equation



The intermediate form, RNH_3OH , is a hydrogen-bonded complex of RNH_2 and water; it may be regarded as a transition state between the two uncharged molecules on the left-hand side of the equation, and the two ions on the right. Many authors in the past, however, have set the concentration of RNH_3OH equal to the total concentration of $(\text{RNH}_2) + (\text{RNH}_3\text{OH})$ and have then formulated their measurements in terms of the constant K_B :

$$K_B = \frac{(\text{RNH}_3^+)(\text{OH}^-)}{(\text{RNH}_3\text{OH})} = \frac{(\text{Conjugate acid})(\text{OH}^-)}{(\text{Total conjugate base})} \quad (5)$$

On the other hand, the convention we have employed here is to denote total conjugate base, $(\text{RNH}_2) + (\text{RNH}_3\text{OH})$, by the symbol (RNH_2) . Experimentally it would be extremely difficult to discriminate between RNH_2 and RNH_3OH , and in practice there is generally no need to do so—we simply lump them together as one single quantity, which *is* readily measured experimentally. Hence when we write the acidity constant, K_A ,

$$K_A = \frac{(\text{RNH}_2)(\text{H}^+)}{(\text{RNH}_3^+)} = \frac{(\text{Total conjugate base})(\text{H}^+)}{(\text{Conjugate acid})} \quad (6)$$

the quantity we have denoted by (RNH_2) in (6) is actually identical experimentally with (RNH_3OH) in (5) above. Moreover (H^+) and (OH^-) are related by the equation $(\text{H}^+)(\text{OH}^-) = K_w$ —see equation (14) below—so that K_B and K_A are related by the equation

$$K_B = \frac{K_w}{K_A} \quad \text{or} \quad pK_B = pK_w - pK_A \quad (7)$$

Thus for ammonia at 25° , $pK_A = 9.3$ and $pK_B = 4.7$. The K_B values found in the literature may readily be converted to K_A values by equation (7). We shall formulate all acid-base equilibria in terms of K_A values in the following discussion.

G. N. Lewis has proposed a very general theory of acids and bases, even more general in character than that of Brönsted and Lowry. According to Lewis' definition, a substance such as boron trichloride, which reacts readily and completely with bases in nonaqueous solvents:



is to be considered as a very strong acid, although it contains no hydrogen. It certainly cannot be regarded as a proton donor, but is an electron acceptor, whereas the base NH_3 is an electron donor. For certain purposes this very broad conception of acids and bases is extremely valuable. In biological systems, however, the ubiquitous presence of water, and of other proton acids, provides an ample reserve of protons available for exchange with bases. Hence for the purposes of the biochemist, Brönsted's

definition of acids as proton donors, and bases as proton acceptors, is sufficiently general.

An interesting discussion of the various meanings which chemists have given to the word "base" is given by W. M. Clark in "Topics in Physical Chemistry" (2nd ed., p. 242).

The strength of an acid may be expressed by its tendency to give off protons, according to the equation



From the law of mass action, the equilibrium constant of this reaction is, in terms of the activities of H^+ , acid, and base,

$$K_A = \frac{a_{\text{H}} a_{\text{R}}}{a_{\text{HR}}} = \frac{a_{\text{H}} a_{\text{base}}}{a_{\text{acid}}} \quad (9)$$

For the activities of the acid and its conjugate base, however, we may substitute the respective concentrations, provided we are considering reactions which take place in a nearly constant medium, in which the temperature and the chemical composition of the solvent do not vary significantly during the reactions under consideration. These conditions are very nearly fulfilled in many systems of biological importance, such as sea water, blood, lymph, and perhaps in the interior of certain tissues. We may, therefore, provisionally substitute concentrations for activities, although this substitution is justifiable only under strictly limited conditions. The more exact formulation, in terms of activities, is undertaken later in this chapter.

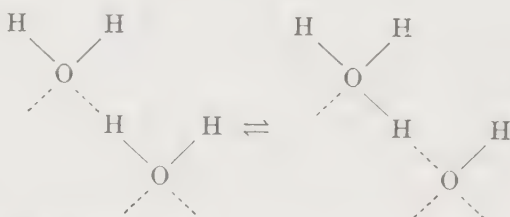
An isolated proton is an exceedingly minute particle. It carries an electric charge and is extremely reactive. It can exist for an appreciable period of time only in a gaseous phase at very low pressures. In any liquid phase, however, a proton reacts, almost as soon as it is liberated, with a neighboring molecule. Thus an acid can act as a proton donor only in the presence of a suitable proton acceptor—that is, another base. Let A_1 denote the acid in question, and B_1 the conjugate base. To a solution of A_1 we now add a base, B_2 , to which the conjugate acid is A_2 . The acid-base exchange reaction is then



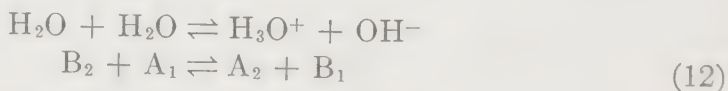
and the equilibrium constant of the reaction is (see equation 9)

$$\frac{(A_2)(B_1)}{(B_2)(A_1)} = \frac{K_{A_1}'}{K_{A_2}'} = K' \quad (11)$$

Here again the quantities in parentheses denote concentrations; later we shall consider the true thermodynamic constants, in which concentrations are replaced by activities. The proton activity has vanished from equation (11), since we are now considering only the ratio of two acidity constants of the type formulated in equation (9). Furthermore reaction (10) is formulated as occurring between the acid-base pairs, A_1 and B_1 , A_2 and B_2 , in a dilute solution in a particular solvent. The ratio K_{A_1}/K_{A_2} is specified only for the conditions existing in that solvent. If we study the same acid-base pair in another solvent, the ratio of the K 's will in general be different. Equations (10) and (11) may be illustrated by reactions occurring in pure water. In the water structure in which H_2O molecules are linked by hydrogen bonds (Chapter 2), exchanges of protons between water molecules must be proceeding constantly; the process may be visualized as the jumping of a proton from the oxygen to which it is attached by a covalent bond to an adjoining oxygen to which it is attached by a hydrogen bond:



This process may be formulated more compactly by an equation which is a special example of the general acid-base equation (10), the water molecules acting as both acids and bases:



This reaction involves a large increase in electrostatic energy, since an anion and a cation are produced from two neutral molecules; consequently the H_3O^+ ion has a strong tendency to donate a proton, either to the OH^- ion or to one of the other oxygens linked to it by hydrogen bonding. Likewise the OH^- ion has a strong tendency to accept a proton from one of its neighbors. At equilibrium the concentrations of H_3O^+ and OH^- are very small; in pure water they are necessarily equal. The amounts of these ions present in pure water have been determined by conductivity measurements and in other ways, to be discussed later. At 25° it has thus been found that $(H_3O^+) = (OH^-) = 1.00 \times 10^{-7}$, to a high degree of precision. The equilibrium constant of reaction (12) may be written

$$\frac{(H_3O^+)(OH^-)}{(H_2O)^2} = K \quad (13)$$

Since, however, in dilute aqueous solutions the concentration of water is very large and remains practically constant throughout any reaction considered, it is conventional to include the factor $(\text{H}_2\text{O})^2$ in the ionization constant of water, K_w , as it is ordinarily written:

$$(\text{H}_3\text{O}^+)(\text{OH}^-) = K(\text{H}_2\text{O})^2 = K_w = 10^{-14.00} \text{ at } 25^\circ \quad (14)$$

It is customary to denote the negative logarithm of an equilibrium constant, K , by the symbol pK ; so we may write $pK_w = 14.00$ at 25° . The value of pK_w decreases—that is, K_w increases—with rising temperature; for instance, $pK_w = 14.94$ at 0° , and 13.59 at 37° .

The oxonium (or hydronium) ion is the acid conjugate to the base water. Any strong acid, when added to water, is converted almost completely to oxonium ion. In the case of HCl , for instance, the reaction is



Here, again, the reaction follows the pattern of equation (10). Hydrochloric acid is so strong an acid that reaction (15) goes almost completely to the right; in water, HCl behaves for all practical purposes as a completely ionized electrolyte, although in solvents of low dielectric constant it is present largely as undissociated HCl and is found to be a much weaker acid than (for instance) perchloric acid. All acids react with water to give oxonium ion; acids which are strong in water, such as perchloric acid and the hydrogen halides, are virtually completely converted into oxonium ion and their conjugate bases. Weak acids, such as acetic, undergo only a partial conversion:



In writing the equilibrium constant for reactions such as this, it is customary to omit the water involved, for the same reasons indicated in the discussion of equations (13) and (14). Furthermore, the oxonium ion is commonly referred to as “hydrogen ion,” the water molecule to which the proton is attached being ignored.² For solutions rich in water, this leads to no confusion or ambiguity. We shall therefore follow this convention and shall ordinarily denote the hydrogen ion concentration in dilute aqueous solutions by the symbol (H^+) .

The Concept of pH; a Preliminary Statement

For many purposes, the logarithm of (H^+) is a more useful quantity than (H^+) itself. Since we shall deal for the most part with solutions in

² The grounds for the belief that “hydrogen ion” in water is oxonium ion are well set forth by L. P. Hammett, “Solutions of Electrolytes” (2nd ed., 1936), and by the same author in more advanced fashion in his “Physical Organic Chemistry.”

which $(\text{H}^+) < 1$, $\log (\text{H}^+)$ is negative for such solutions and $-\log (\text{H}^+)$ is positive. It is conventional among workers in this field to denote the negative logarithm of a quantity, X , by pX , as we have already done with pK_w ; so we write $-\log (\text{H}^+) \equiv pH$. This is to be regarded as a preliminary definition of pH , in which we are disregarding the distinction between concentrations and activities of the hydrogen ion and other ions in the system. Later we shall consider the problems which arise when the attempt is made to measure the activities of ions and shall formulate a more precise definition of pH in terms of the experimental operations ordinarily employed in measuring it.

The precise establishment of an accurate pH scale is a complicated task. It is easy, however, with our present knowledge, to fix approximately certain points on such a scale. In pure water the hydrogen ion (oxonium ion) concentration must equal the hydroxyl ion concentration. Hence we have, from (14), for water at 25° ,

$$(\text{H}^+)^2 = 10^{-14.00} \quad \text{or} \quad (\text{H}^+) = 10^{-7.00} \quad \text{or} \quad pH = 7.00 \quad (17)$$

A molar solution of HCl in water is, as we have seen, virtually completely ionized. The hydrogen ion *concentration*, therefore, is equal to the total HCl concentration in such a solution. On this basis $(\text{H}^+) = 1$ and $pH = 0$ in this solution. The effect of the electrical attraction between oppositely charged ions is to reduce the activity of all the ions in solution (Chapter 5). Hence the activity of the H^+ ion is slightly less than the concentration. In a $0.1 M$ solution of HCl , $(\text{H}^+) = 0.1$ and $pH = 1$, if interionic forces are assumed to be negligible. The corrections to be introduced later, arising from interionic attraction effects, lead to a somewhat higher pH value, approximately 1.08. In this preliminary survey, however, we shall consider an estimated pH value satisfactory if it is reliable within 0.1 pH unit, so that we disregard this distinction for the present.

Likewise, a solution of a strong base, such as Na^+OH^- , may be taken as completely ionized in water. Thus (OH^-) in a $0.1 M$ solution of such a base in water is very nearly 0.1 and $pOH = -\log (\text{OH}^-)$ is very nearly 1. Since $pH + pOH = 14$, from equation (14), pH in this system is near 13.

Calculations of pH in Systems of Acids of Known Acid Strengths

The preceding discussion gives an approximate indication of the nature of the pH scale in water or dilute salt solutions. Temporarily we defer attempts at more precise definition in order to consider a class of systems of great biological importance, composed of one or more weak acids and their conjugate bases. Let us assume the K_A value of an acid to be known, and calculate the pH as a function of the (acid)/(base)

ratio. Taking the logarithms of the terms in (9), setting $pH = -\log (H^+)$ and $pK_A = -\log K_A$, we obtain

$$pH = pK_A + \log \frac{(\text{base})}{(\text{acid})} = pK_A + \log \frac{\alpha}{1 - \alpha} \quad (18)$$

In equation (18), the symbol α denotes the ratio of the molecules of conjugate base to the total number of molecules of acid plus conjugate base:

$$\alpha = \frac{(\text{base})}{(\text{acid}) + (\text{base})} = \frac{K_A}{(H^+) + K_A} \quad (19)$$

In the analysis of problems concerning acid-base equilibria, it is frequently necessary to calculate (H^+) in systems containing one or more weak acids and their conjugate bases, often in the presence of various other added ions and molecules. If the values of K_A for the acids involved are known, the analysis of the problem may be carried out in a series of steps as follows:

1. Write out explicitly the equations for the various equilibrium constants of the acids in the system, including the value for K_w .
2. Formulate equations indicating what is known of the total stoichiometric concentrations of the various molecular and ionic species involved.

3. The system as a whole must be electrically neutral; therefore the equivalent concentrations of all the anions present must be equal to the equivalent concentration of all the cations. Formulate explicitly the equation of electrical neutrality in terms of all the ions present.

4. Consider which ions present are likely to be of negligible concentration relative to the others, and eliminate them from the equation of electrical neutrality. This step requires some judgment, which is acquired only by a certain amount of experience in dealing with these problems. Generally the validity of the assumptions made can be checked later, after a preliminary trial solution of the problem in which certain factors have been neglected.

We may give a first illustration of these points in terms of a simple system. The value of K_A for acetic acid has been found to be $10^{-4.75}$. What are the values of (H^+) and of pH in a solution of acetic acid in water at concentration C ?

1. Two dissociation constants are involved:

$$K_A (\text{acetic acid}) = \frac{(H^+)(A^-)}{(HA)} = 10^{-4.75} \quad (20)$$

and

$$K_w = (H^+)(OH^-) = 10^{-14} \text{ at } 25^\circ \quad (21)$$

2. The total concentration of acetic acid is C ; we therefore have the relation

$$(\text{HA}) + (\text{A}^-) = C \quad (22)$$

since the added acid must be present in one of these two forms.

3. Only two kinds of anions are present, and one kind of cation. The equation of electrical neutrality is

$$(\text{H}^+) = (\text{A}^-) + (\text{OH}^-) \quad (23)$$

4. We know from experience that a dilute solution of acetic acid is distinctly acid to neutrality; thus we may assume that $(\text{OH}^-) \ll 10^{-7}$. We therefore disregard (OH^-) in equation (23), which now becomes $(\text{H}^+) = (\text{A}^-)$. Combining (22) and (23) we may then write

$$K_A = \frac{(\text{H}^+)(\text{A}^-)}{(\text{HA})} = \frac{(\text{H}^+)^2}{C - (\text{H}^+)} = 10^{-4.75} \quad (24)$$

This may be written in the standard form of a quadratic equation. With $(\text{H}^+) = X$ for convenience, (24) becomes

$$X^2 + K_A X - K_A C = 0 \quad (25)$$

and

$$2X = 2(\text{H}^+) = -K_A + [K_A^2 + 4(K_A C)]^{1/2} \quad (26)$$

Actually a still simpler solution is often possible if K_A is sufficiently small and C sufficiently large. In that case we may treat (H^+) in the denominator of equation (24) as negligible in comparison with C , and the equation then becomes

$$\begin{aligned} (\text{H}^+)^2 &= K_A C \\ \text{pH} &= \frac{\text{p}K_A - \log C}{2} \end{aligned} \quad (27)$$

For 0.1 M acetic acid, for instance, we find the pH , by equation (26), to be 2.878 to the third decimal place; by equation (27), it is 2.875. Here the simple equation is certainly adequate, if we wish to calculate pH only within ± 0.01 . If we were dealing with more dilute solutions, or with a stronger acid, the error would be larger, and it might be necessary to use the more complete equation. The error involved in setting $(\text{OH}^-) = 0$ in equation (23) is certainly negligible here; but this might not be true if we were working at very great dilutions.

Equation (24) applies to a cationic acid, such as a solution of ammonium chloride at the concentration C , provided we write (NH_4^+) instead of (HA) , (NH_3) instead of (A^-) , and $K_A = 10^{-9.3}$ instead of $10^{-4.75}$. Here however, the assumption that (OH^-) is negligible, though justified, is not so obvious as in the case of acetic acid.

The Titration of a Weak Acid with a Strong Base: Buffer Action

If a very strong base, such as hydroxyl ion—say in the form of Na^+OH^- —is added to a solution of a weak acid, HA , the reaction



runs toward the right, virtually to completion. In a solution to which C moles of acetic acid and X moles of Na^+OH^- have been added, per liter, we may calculate the pH value by the same principles as before, if $X < C$. The two equilibrium constants, (20) and (21), are still the only ones required, but now an additional cation—the sodium ion—is present. Since it is not involved in the reaction formulated above, its concentration (m) is the same as that of the added sodium hydroxide. The equation of electrical neutrality then becomes

$$(\text{Na}^+) + (\text{H}^+) = (\text{A}^-) + (\text{OH}^-) = m + (\text{H}^+) \quad (28)$$

The stoichiometry of equation (22) holds as before.

Here, again, we try simplifying (28) by treating (OH^-) as negligible. This is not so clearly justifiable as in the case of acetic acid alone in water, but we can take it as known from experience that the solution remains acid throughout the titration until X is nearly equal to C . Then we have for the concentration of acetate ion (A^-) the presumably very close approximation

$$(\text{A}^-) = (\text{Na}^+) + (\text{H}^+) = X + (\text{H}^+) \quad (29)$$

Noting that $\alpha = [X + (\text{H}^+)]/C$, we may write

$$K_A = \frac{(\text{H}^+)(\text{A}^-)}{(\text{HA})} = \frac{(\text{H}^+)[X + (\text{H}^+)]}{[C - X - (\text{H}^+)]} = \frac{(\text{H}^+)\alpha}{1 - \alpha} \quad (30)$$

Here α is defined by (19). In logarithmic form

$$\text{pH} = \text{p}K_A + \log \frac{[X + (\text{H}^+)]}{[C - X - (\text{H}^+)]} = \text{p}K_A + \log \frac{\alpha}{1 - \alpha} \quad (31)$$

Equation (30) as it stands can be solved as a quadratic. It is simpler, however, to make use of the fact that (H^+) is in general much smaller than either C or X . Thus we may temporarily disregard (H^+) in the two expressions in brackets in the numerator and denominator of (30), obtaining the simplified relation

$$K_A = \frac{(\text{H}^+)(X)}{C - X} = \frac{(\text{H}^+)\alpha}{1 - \alpha} \quad (32)$$

$$\begin{aligned} \text{pH} &= \text{p}K_A + \log \left(C \cdot \frac{X}{C - X} \right) = \text{p}K_A + \log \frac{\alpha}{1 - \alpha} \\ &= \text{p}K_A + \log \frac{(\text{base})}{(\text{conjugate acid})} \end{aligned} \quad (33)$$

A preliminary value of (H^+) derived from (32) may now be inserted into the terms in brackets in (30) or (31) and the resulting equation solved to obtain a more accurate value of (H^+) . If necessary, this value can again be inserted in the brackets to obtain a still further approximation for (H^+) , but such a refinement is rarely necessary. For many purposes, indeed, in solutions of an acid of the strength of acetic acid, it may be quite accurate enough to use equation (32) without further correction—that is, to set X equal to the sodium ion concentration, and $C - X$ equal to the total acid originally present, minus the added sodium ion. If the pK value of the acid is higher—say 6 or above—the error involved in using equation (32) instead of (31) is quite negligible, provided X and $C - X$ are of the order of 0.01 M or more. This is the approximation originally introduced by L. J. Henderson, in his development of the theory of buffer action. Equation (33) is commonly known as the Henderson-Hasselbalch equation.³ We may use it to calculate the change of pH occurring during the titration of a weak acid with a strong base. It is convenient to make the calculations on the assumption that the volume of the system remains constant during the titration, and it is often possible to carry out the titration so that this is approximately true. Then the total concentration $C = (HA) + (A^-)$ also remains unaltered during the titration. If X moles of OH^- ion have been added, per mole of HA originally present, we take $(A^-) = X$ according to the approximation of equation (32), and for an increment ΔX in X the corresponding increment in α is

$$\Delta\alpha = \frac{\Delta X}{C} \quad (34)$$

The form of the resulting titration curve, when pH is plotted against the added alkali X , is shown in Fig. 1. The form of the curve, on this simple theory, is independent of the pK_A value of the particular acid and depends only on α . Two acids with different pK_A values thus give titration curves which can be superposed simply by a horizontal displacement of one curve along the abscissa axis until it coincides with the other (see Fig. 1). In a constant solvent medium, with the ionic strength maintained constant, or nearly so, throughout the titration, equation (33) describes the

³ In the past it has generally been conventional to write (salt) instead of (base) in equation (33). Thus in a system containing acetic acid and sodium acetate the base present is acetate ion. Since sodium acetate is completely dissociated, and acetic acid generally very little dissociated, the stoichiometric concentration of the salt sodium acetate is nearly equal numerically to that of acetate ion. The advantage of the formulation given in equations (18) and (31) or (33), however, is that it applies to all types of acid-base pairs, such as those listed in Table I, whereas the traditional formulation in terms of salt concentration does not apply, for instance, to the equilibrium between ammonia and ammonium ion.

titration curves of many weak acids with high precision, whether they are uncharged acids such as acetic acid, or cationic acids such as the ammonium ion. The effects of variation in ionic strength or dielectric constant on the value of pK_A are considered later.

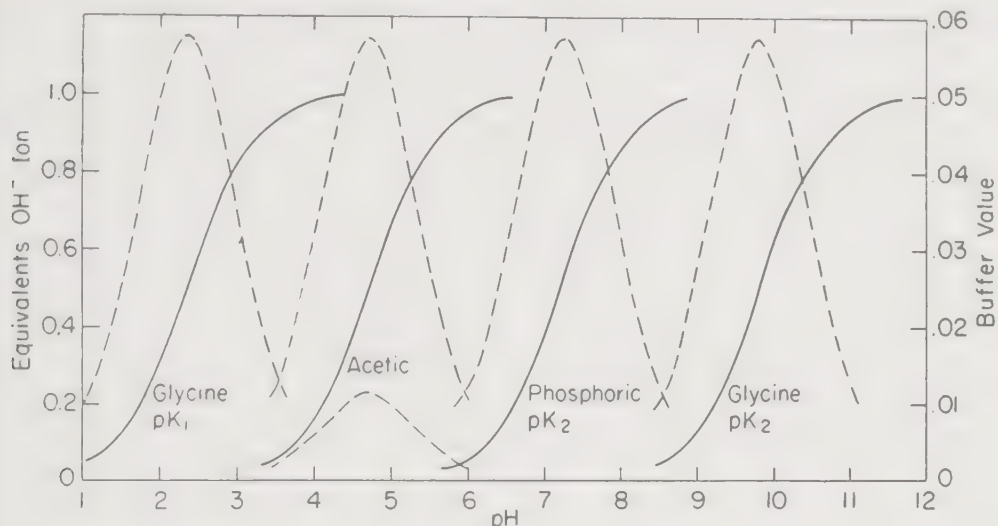


FIG. 1. Titration curves of glycine (pK_1), acetic acid, phosphoric acid (pK_2), and glycine (pK_2) are illustrated by the solid lines. Corresponding buffer values for 0.1 *M* solutions are indicated by the dotted lines. For acetic acid the buffer value of a 0.02 *M* solution is also indicated. Note that the *pH* of maximum buffer value for any given acid coincides with the pK value of that acid.

The characteristic feature of such curves as those shown in Fig. 1 is that when α is near 0.5—in other words when *pH* is nearly equal to pK_A —the addition of a relatively large amount of strong acid or base produces little change in *pH*. This phenomenon, known as buffer action, is of the most fundamental importance for the stability of the living organism. Small alterations in *pH* often profoundly alter the rates of many biochemical reactions, and a shift of a few tenths of a unit in the *pH* of blood may be an index of the difference between an organism in normal health and one on the point of death. The maintenance of a nearly constant *pH* value in systems to which relatively large amounts of acids or bases may be added is a stabilizing factor of major importance. Acids such as lactic or pyruvic acid, for instance, are frequently products of metabolic reactions, and the fluids of the living organism must be equipped to receive considerable increments of such substances without undergoing major fluctuations of *pH*.

It is convenient to have a quantitative expression to describe the resistance of a system to *pH* changes brought about by the addition of strong acid or alkali. The obvious quantity to employ is the slope of the

titration curve at any point, when X is plotted⁴ against pH , or, in differential notation, dX/dpH . This is the "buffer value" of van Slyke. We may evaluate it as follows: Equation (33) gives pH as a function of α , and (34) gives α as a function of X . Hence we may write

$$\frac{dX}{dpH} = \left(\frac{dX}{d\alpha} \right) \left(\frac{d\alpha}{dpH} \right) = C \frac{d\alpha}{dpH} \quad (35)$$

the second equality being derived from (32). On differentiating (18) we obtain

$$dpH = d \log \left(\frac{\alpha}{1 - \alpha} \right) = \frac{1}{2.303} d \ln \left(\frac{\alpha}{1 - \alpha} \right) = \frac{d\alpha}{2.303\alpha(1 - \alpha)}$$

and

$$\frac{d\alpha}{dpH} = 2.303\alpha(1 - \alpha) \quad (36)$$

Hence, the buffer value is, from (35) and (36),

$$\frac{dX}{dpH} = 2.303C\alpha(1 - \alpha) \quad (37)$$

This equation is valid provided the concentrations of the base A^- and the conjugate acid HA are large in comparison with both the hydrogen and hydroxyl ion concentrations. If (H^+) or (OH^-) is considerable, additional terms are required, which, however, are simply derived (see, for instance, Clark, 1928, Chapter II, or Bates, 1954). Between pH 4 and 10, and therefore in all systems of physiological interest, equation (37) can be safely applied.

It is plain that the buffer value of a system containing an acid and its conjugate base is proportional to the total concentration, C . As α approaches either zero or unity, the buffer value falls toward zero; its maximum value is attained when $\alpha = 1 - \alpha = 0.5$ and is equal to $0.575C$. The buffer value for a system containing several different pairs of conjugate acid and base components is the sum of the buffer values for the individual component systems. Buffer values for mixtures of several acids and their conjugate bases are shown by the dashed lines in Fig. 1.

The Experimental Determination of Acidity Constants (K_A Values)

The nature of acid-base equilibria cannot be fully understood without some consideration of the experimental methods used to determine them.

⁴ The addition of strong acid to a buffer mixture may be considered mathematically simply as a negative increment in X .

We shall consider three principal methods, involving widely different experimental procedures.

THE CONDUCTIVITY METHOD

This method, in a simple form, was widely applied by Wilhelm Ostwald and his school, beginning about 1890, and later by many others. It has been greatly refined in recent years, to yield results of very high accuracy. It is well adapted for the study of solutions of pure acids and bases, dissolved in water or other solvents; but it is not suitable for the study of buffer mixtures, or of systems containing added neutral salts. We shall first state the method in a relatively simple form, essentially that used by Ostwald, and then indicate the nature of some of the modern refinements.

Consider the reaction $\text{HA} + \text{H}_2\text{O} = \text{A}^- + \text{H}_3\text{O}^+$, in which a neutral acid reacts with water to form a negative conjugate base and the hydronium ion. The electrical conductance of the solution is due to these ions; the contribution of the ions derived from the water itself is generally negligible by comparison, but can be allowed for if necessary. If we neglect this, we may set $(\text{A}^-) = (\text{H}_3\text{O}^+) = \alpha C$, where C is the total stoichiometric concentration of added acid, and α (often called the "degree of dissociation") is the fraction of this total which is in the form of the conjugate base. The concentration of HA is obviously equal to $(1 - \alpha)C$, and the acid dissociation constant is (writing H^+ for H_3O^+)

$$K_A = \frac{(\text{H}^+)(\text{A}^-)}{(\text{HA})} = \frac{\alpha^2 C}{(1 - \alpha)} \quad (38)$$

If we could neglect interionic forces, and if α were equal to unity, the equivalent electrical conductance would simply be equal to the limiting value, Λ_0 , for the sum of the H_3O^+ ion and the A^- ion. This quantity is accessible to experimental measurement, since in general the sodium or potassium salt, Na^+A^- or K^+A^- , is a completely dissociated strong electrolyte. Thus Λ_0 for these salts may be determined experimentally by the methods outlined in Chapter 7; and the same may be done for a strong acid such as H^+Cl^- , and for a salt with a common anion, such as Na^+Cl^- . Then it follows from Kohlrausch's law of the independent migration of the ions that

$$\Lambda_0(\text{H}^+\text{A}^-) = \Lambda_0(\text{Na}^+\text{A}^-) + \Lambda_0(\text{H}^+\text{Cl}^-) - \Lambda_0(\text{Na}^+\text{Cl}^-) \quad (39)$$

In the actual solution of HA , however, the equivalent conductance is not Λ_0 , since only the fraction α of the acid HA has been converted into the ions. Neglecting all the complications due to interionic forces, we should then naturally suppose that the measured equivalent conductance (Λ)

would be simply $\alpha\Lambda_0$. This is indeed what Ostwald assumed, setting $\alpha = \Lambda/\Lambda_0$ in equation (39), and determining the equivalent conductance at a series of values of the concentration C . The values of K so calculated from data at different values of C were reasonably constant and have served to provide valuable material for approximate computations of the relative acidities of different substances.

To obtain really accurate values of K_A from such measurements, however, we must make two kinds of corrections for interionic forces.

1. The ionic strengths of dilute solutions of weak acids are by no means negligible. In a solution of acetic acid ($K = 10^{-4.75}$) at 0.01 M , for example, one may calculate from equation 38 that $\alpha = 0.045$ approximately. For such a solution of a weak uni-univalent electrolyte the ionic strength $\omega = \alpha C = 0.01 \times 0.045 = 0.00045$, and $\sqrt{\omega}$ is equal to 0.021. At this ionic strength we may employ the limiting equation of the Debye-Hückel theory for aqueous solutions (Chapter 5, equation 76.3) in the form

$$\log f_{\pm} = -0.5 \sqrt{\omega} = 0.0105$$

The corresponding value of f_{\pm} is 0.976, which deviates significantly from unity. Taking account of activity coefficients,* therefore, we may rewrite equation (38) in the form

$$K_A = \frac{(a_H)(a_{A^-})}{(a_{HA})} = \frac{f_H f_{A^-}}{f_{HA}} \frac{(H^+)(A^-)}{(HA)} = \frac{f_{\pm}^2 (H^+)(A^-)}{f_{HA} (HA)} \quad (40)$$

It is customary to consider f_{HA} as unity at low ionic strengths, since HA is uncharged and the Debye-Hückel term therefore vanishes for it. The assumption is not quite exact but works well in practice, for reasons which we consider in more detail later. Thus, since f_{\pm} is less than unity, the *concentrations* of H^+ and A^- ions must be greater than they would be if they were calculated from the limiting value of the thermodynamic constant, K_A , with no correction for activity coefficients. Failure to correct for activity coefficients would thus yield an apparent value of K which is larger than the true K .

2. Owing to interionic forces the mobility of the ions is less than it would be at infinite dilution. The magnitude of the deviations to be expected is given by Onsager's equation (Chapter 7, equation 24). The observed value of Λ , at finite concentration, is *less* than $\alpha\Lambda_0$, and the calculated degree of dissociation is therefore too small. The error due to neglect of the difference between Λ and Λ_0 is thus in the opposite direction

* As in Chapter 5, we here use the symbol f to denote an activity coefficient (ratio of activity to mole fraction). Later in this chapter, however, we employ the symbol γ , with appropriate subscript, to denote the activity coefficient, defined as the ratio of activity to molality or to molar concentration.

from that due to neglect of activity coefficients. Owing to this partial cancelation of errors, the values of K_A obtained by Ostwald and his school were generally not very far from the true values, often within 0.1 in pK_A . The best modern work, however, such as that of D. A. MacInnes and T. Shedlovsky, has yielded pK values, extrapolated to zero ionic strength, which are reliable within ± 0.002 and sometimes within ± 0.001 , as shown by their agreement with electromotive force values, determined by methods discussed below. Conductivity measurements on solutions of weak bases, such as ammonia, can be used to determine the basic dissociation constant, K_B (equation 5); the reasoning involved is exactly analogous to that given here for acids such as acetic acid.

THE INDICATOR METHOD FOR DETERMINING RELATIVE STRENGTHS OF ACIDS

The indicator method is derived from the observation that many acids and bases are colored, and that in such cases the color of the acid is different from that of its conjugate base. The term "color," of course, may be used very broadly—the essential requirement is that the acid and the conjugate base differ in their absorption of light at some particular wavelength; the region chosen for study may lie in the ultraviolet, the visible, or the infrared. For accurate work, however, the light should be monochromatic, or nearly so, in order to define quantitatively the relation between light absorption and concentration for the acid and its conjugate base.

The use of the method generally involves the interaction of two acid-base pairs—we may denote them as A_1 and B_1 , and as A_2 and B_2 , respectively. The observations are carried out with light of a given wavelength λ , so chosen that the members of one pair—say A_2 and B_2 —do not absorb the light appreciably, whereas either A_1 or B_1 absorbs it to an extent which is readily measurable. Commonly both absorb it, but the absorption (or extinction) coefficient must of course be different for A_1 and B_1 if the method is to be used.

The experiment may be carried out, for instance, by adding the acid A_1 , at concentration S_1 , to the base B_2 , at concentration S_2 . They react according to equation (10), forming A_2 and B_1 in equivalent amounts. We may write X for the equilibrium concentration of each:

$$X = (B_1) = (A_2)$$

Then the concentration of A_1 at equilibrium is $S_1 - X$, and that of B_2 is $S_2 - X$. Hence the equilibrium constant, K' , of equation (11) is

$$K' = \frac{(A_2)(B_1)}{(B_2)(A_1)} = \frac{X^2}{(S_1 - X)(S_2 - X)} = \frac{K_{A_1}'}{K_{A_2}'} \quad (41)$$

Moreover we note that the ratio $(B_1)/(A_1) = X/(S_1 - X)$ is equal to the ratio $\alpha/(1 - \alpha)$ of base to conjugate acid, as defined in equations (30) to (33), inclusive. Let ϵ_{A_1} be the molar extinction coefficient⁵ of A_1 for light of wavelength λ , and ϵ_{B_1} the corresponding extinction coefficient of B_1 . These extinction coefficients may be determined spectrophotometrically in separate experiments on solutions of pure A_1 and pure B_1 at known concentrations. In the mixed solution at equilibrium, as defined by (41), the molar extinction coefficient, ϵ , is determined per mole of A_1 plus B_1 . One mole of $A_1 + B_1$ contains $1 - \alpha$ moles of A_1 and α moles of B_1 . Hence

$$\epsilon = (1 - \alpha)\epsilon_{A_1} + \alpha\epsilon_{B_1} \quad (42)$$

and therefore

$$\frac{\alpha}{1 - \alpha} = \frac{X}{S_1 - X} = \frac{(B_1)}{(A_1)} = \frac{\epsilon - \epsilon_{A_1}}{\epsilon_{B_1} - \epsilon} \quad (43)$$

Thus from (43) it is possible to evaluate X , which is the only unknown in (41), and thereby to determine the ratio K_{A_1}'/K_{A_2}' . This is the basis of the indicator method for determining the relative strengths of different acids. It is a powerful method, widely applicable in aqueous solutions and in most nonaqueous media.

Several comments may be made on the use of the indicator method:

1. Light absorption measurements give a measure of the *concentration* of the molecular species under study, not of its activity. Thus the equilibrium constant, K' , in (41) is given by such measurements directly in terms of concentrations. If the solvent medium is varied, for instance by varying the ionic strength or by adding a new component to the solvent which raises or lowers its dielectric constant, the activity coefficients of A_1 , A_2 , B_1 , and B_2 may be expected to vary. Hence K' is only an "apparent constant," depending on the particular medium employed and to some extent on the concentrations of the reactants and products. By working at various ionic strengths, however, in an otherwise constant medium, and extrapolating the values of K' , in a suitable fashion, to zero ionic strength, the true limiting value of K at zero ionic strength may be determined. The nature of the extrapolations involved is discussed later, in connection with electromotive force methods.

⁵ If light of wavelength λ traverses a distance d in the solution, and the light intensity is thereby reduced from I_0 to I_d , then the molar extinction coefficient, ϵ_A , of an absorbing substance, A , is given by the relation

$$(1/d) \log (I_0/I_d) = \epsilon_A (A)$$

If the substance obeys Beer's law, then ϵ_A is independent of the concentration (A) over the range in which we are interested. This condition is essential for the success of the method.

2. In order to obtain accurate results, the indicator acid, A_1 , and its conjugate base, B_1 , should be present at equilibrium in a ratio not very far from unity; in other words the ratio X/S should not be too close to either zero or unity. This may be seen directly by examination of equation (43), from which it is also apparent that the difference $\epsilon_{A_1} - \epsilon_{B_1}$ should be large if precise measurements are to be obtained. It is, of course, convenient if one of these two coefficients is zero—as, for instance, in the case of phenolphthalein, only the basic form of which absorbs visible light.

Thus, to obtain accurate measurements, the value of K_A for the indicator acid should not be very far from K_{A_2} of the other acid under study; otherwise X will be too large or too small. If, however, K_{A_1} is somewhat—but not much—greater than K_{A_2} , then a similar indicator study may be made with another acid, A_3 , with an acid dissociation constant, K_{A_3} , somewhat less than K_{A_1} . Then the ratio K_{A_2}/K_{A_1} is immediately given, since both are known relative to K_{A_3} . The use of other acid-base pairs, including other indicators, thus permits us to map out a scale of relative K_A values for a series of acids of widely differing strength, in any given solvent medium. All these are relative constants; if the absolute value of any one of them can be fixed by any suitable experimental procedure, or if it is simply defined by an arbitrary convention, then all the other K 's are fixed by reference to that standard.

3. Other techniques than light absorption may be used as the basis of the indicator method. For instance, the Raman frequencies, which are obtained by observing the scattered light from the substance under study, using monochromatic exciting light, are highly characteristic of the molecular structure. If particular Raman frequencies, characteristic of an acid and of its conjugate base, can be quantitatively determined, then the Raman spectrum can be used as the basis of an indicator technique.

Indeed an indicator method need not be optical at all. If acid and conjugate base differ with respect to their magnetic properties, or in any other physical properties which can be measured with sufficient accuracy, then the measurement of such a property may serve as the basis for a quantitative study of acid-base equilibria.

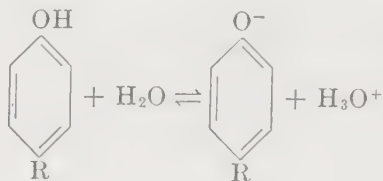
4. If electrometric pH measurements are carried out on the solution, by the techniques discussed later (p. 447), together with the light absorption measurements, then we may combine the results of the two methods by equations (33) and (43), giving

$$pH = pK_A' + \log \frac{\alpha}{1 - \alpha} = pK_A' + \log \frac{\epsilon - \epsilon_A}{\epsilon_B - \epsilon} \quad (43a)$$

Here we have used the generalized subscripts A and B to denote any acid-base pair which differ in light absorption. This is particularly helpful

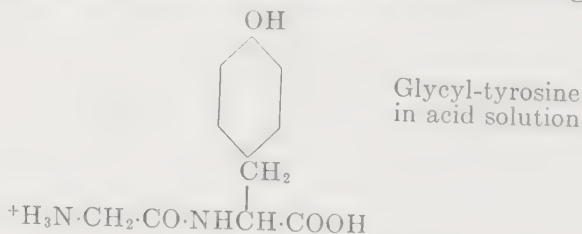
in the study of polyvalent acids, in which one or more acidic groups, or their conjugate basic groups, absorb light of wavelength λ , whereas the other groups in the molecule do not. It is then possible to follow the state of ionization of this particular set of groups as a function of pH .

Consider for instance an ionization process involving a phenol derivative:



The un-ionized phenol shows light absorption with a maximum near 2750 Å, whereas its conjugate base shows maximum absorption near 2930 Å, with a value of ϵ considerably greater than that of the un-ionized phenolic group at this wavelength. In the amino acid tyrosine, for instance, R is $-\text{CH}_2\text{CH}(\text{NH}_3^+)\text{COO}^-$. The pK_A value of the phenolic group is very close to that of the ammonium group, and in an ordinary pH titration the two overlap to a large extent. Neither the ammonium group, however, nor the amino group which is its conjugate base, absorbs light in the region between 2800 and 3000 Å. Hence by measuring the ultraviolet absorption at a suitable wavelength as a function of pH it is possible to determine specifically the fraction of the phenolic hydroxyl groups present in solution which are ionized at any given stage of the titration. The same technique is applicable to proteins, which generally contain many such groups per molecule; interesting and important results have been obtained from such measurements (see Chapter 9). In a protein, however, the absorption maxima are commonly slightly different, both in position and in intensity, from what they are in the corresponding groups in free amino acids, and this factor must be allowed for.

The change of phenolic absorption spectra with pH is illustrated in Fig. 2, which shows the spectrum of glycyl-L-tyrosine as a function of pH . This molecule, like tyrosine, contains three ionizing groups—a car-



boxyl, an amino, and a phenolic hydroxyl group. The pK' value of the carboxyl group is very nearly 3 (Greenstein, 1932); hence the curve shown for pH 1.09 is for the molecule with an unionized carboxyl group, as in

the formula shown. When the carboxyl group ionizes (see curve for pH 4.6) there is very little change in the spectrum; the absorption curve is slightly depressed in the region below $250\text{ m}\mu$, and is slightly raised and displaced to longer wavelengths in the region from about 270 to $290\text{ m}\mu$. The $-\text{NH}_3^+$ group is the next to ionize, with a pK' value of 8.25 at 25° and ionic strength 0.16 . The change in absorption spectrum due to the loss of a proton from the amino group is also small; in O-methyltyrosine,

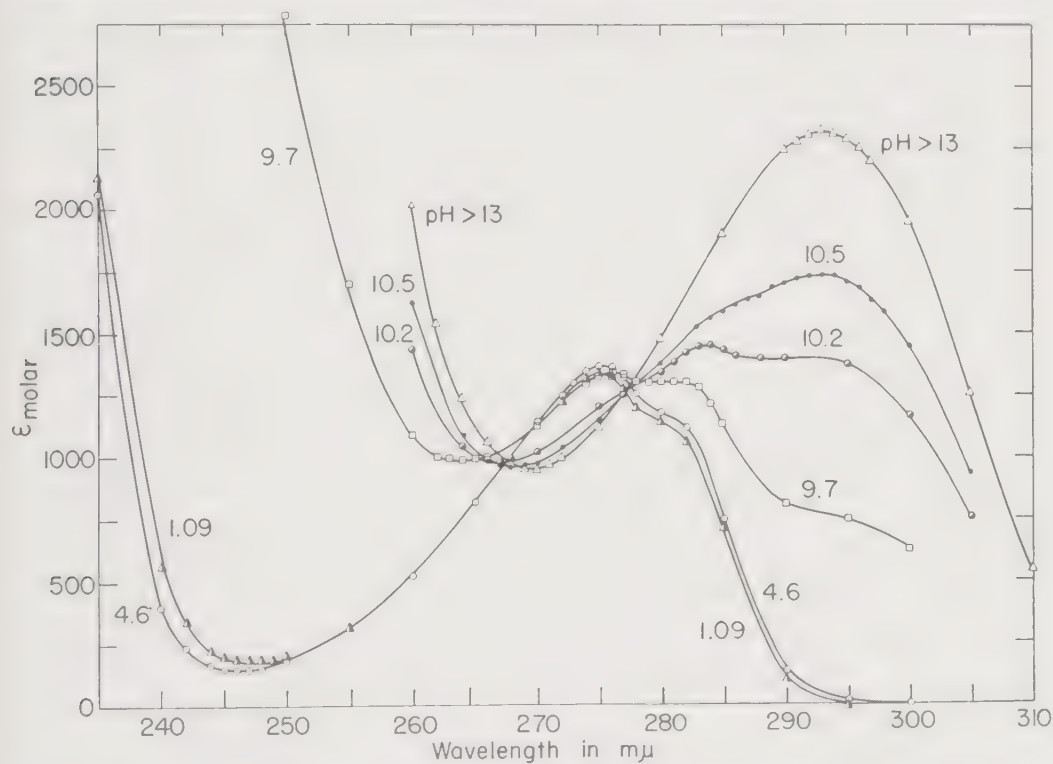


FIG. 2. Ultraviolet absorption spectrum of glycyl-L-tyrosine at various pH values. Temperature 25° , ionic strength 0.16 . (Data of Barbara R. Hollingworth.)

in which the hydroxyl group is blocked and cannot ionize, the change from an $-\text{NH}_3^+$ to an NH_2 group produces a further small displacement of the curve to longer wavelengths, in the range from 270 to $290\text{ m}\mu$. In this, as in other tyrosine derivatives that have been studied, ϵ falls practically to zero at $295\text{ m}\mu$ and higher wavelengths. However, when the phenolic hydroxyl group ionizes, as shown by the progressive changes in the four upper curves at the right of Fig. 2, there is a profound change of absorption; the position of the maximum in ϵ_{molar} shifts from near 275 to $293\text{ m}\mu$, and the value of ϵ at the maximum increases from 1375 to 2325 . At pH above 13 (curve at upper right) the hydroxyl group is completely ionized, and we can take the value of ϵ at the high pH as the value for ϵ_B in (43a). If we make calculations from data at 295 or $300\text{ m}\mu$, ϵ_A ,

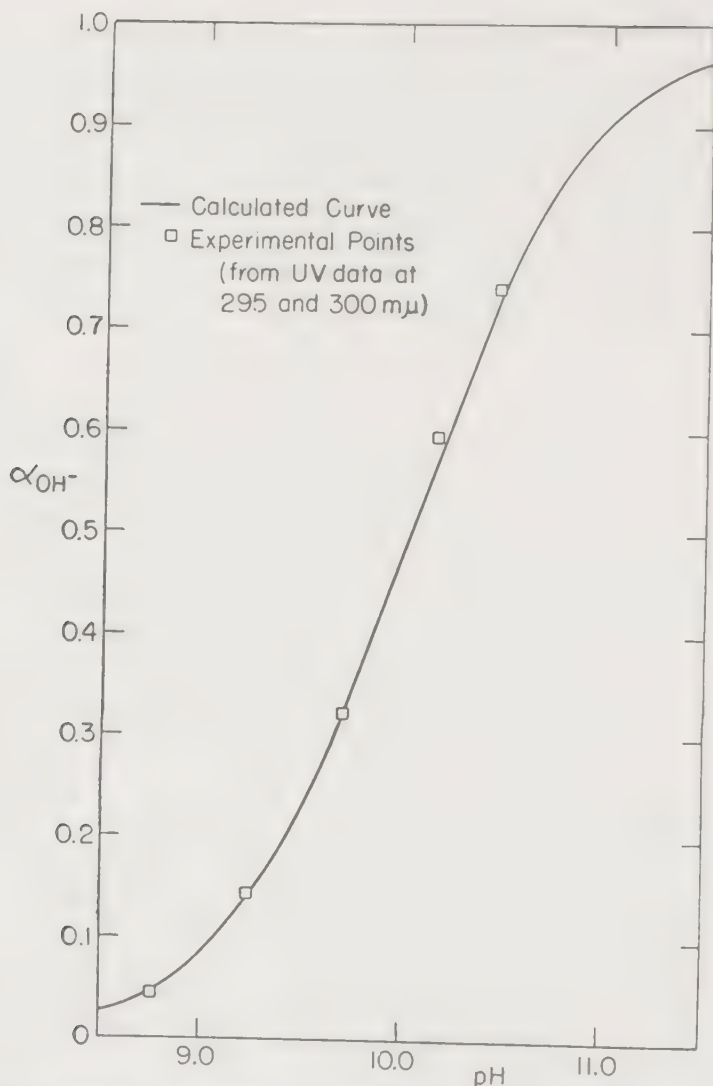


FIG. 3. The fractional ionization (α) of the phenolic hydroxyl group in glycyl-L-tyrosine, calculated from the data of Fig. 2 and other ultraviolet absorption measurements.

as given by the curve at pH 4.6, is virtually zero at both these wavelengths. Thus if the pH and the value of ϵ are both measured for a given solution, we have from (43a):

$$pH = pK_A' + \log [\alpha/(1 - \alpha)] = pK_A' + \log [\epsilon/(\epsilon_B - \epsilon)]$$

The value of α for the phenolic hydroxyl group ($\alpha_{OH} = \epsilon/\epsilon_B$) is shown as a function of pH in Fig. 3. The points are experimental data, from absorption measurements at 295 and at 300 $m\mu$ (identical values of α were obtained at any given pH for both these wavelengths) and the curve is

calculated, assuming $pK_A' = 10.03$ for the hydroxyl group in glycylytyrosine. The data fit the calculations satisfactorily.

For comparison, a pH titration of isoelectric glycylytyrosine is shown in Fig. 4. The ordinate gives Z , the mean net charge on the glycylytyrosine molecules, which is equal to the moles of OH^- ion added per mole of glycylytyrosine, except in the most alkaline solutions where a correction must be made for the free OH^- ion remaining in solution. The calculated curve is based on the values of pK_A' given above (8.25 for pK_2' and 10.03

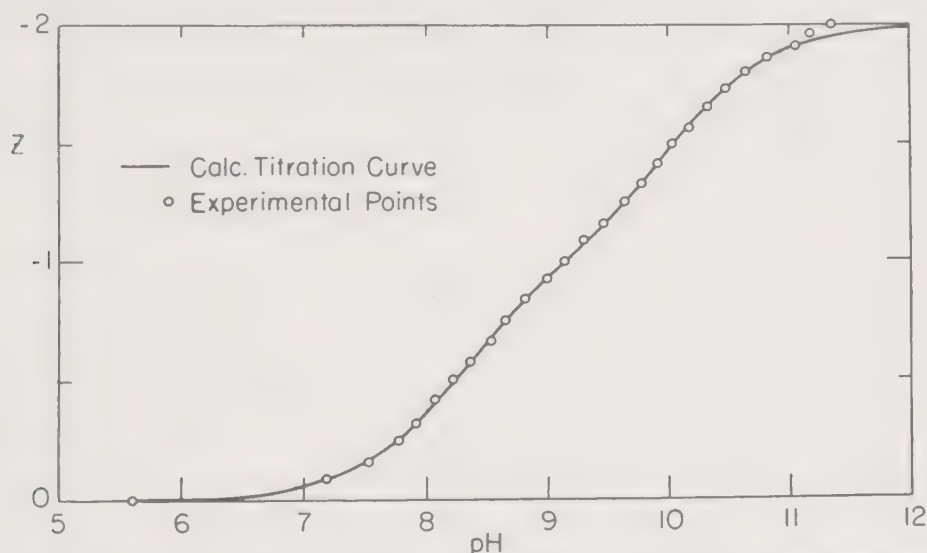


FIG. 4. pH titration of the amino and phenolic hydroxyl groups of glycyly-L-tyrosine at 25° and 0.16 ionic strength. (Data of Barbara R. Hollingworth.)

for pK_3'). Here again the fit is seen to be good. The same value of pK_3' is obtained by measuring ϵ as a function of pH as by measuring OH^- added as a function of pH , for in this case the dissociation of the $-NH_3^+$ group is fairly widely separated from that of the phenolic $-OH$ group. For tyrosine itself, and for most proteins, there is marked overlapping between the ionization of amino groups and phenolic $-OH$ groups, and the determination of α for the $-OH$ groups by spectrophotometric titration is then particularly valuable.

ELECTROMOTIVE FORCE MEASUREMENTS AND pH DETERMINATION

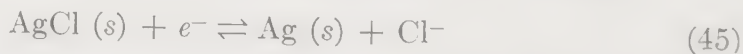
The most powerful and most widely used method for the determination of acid dissociation constants and of pH is the electromotive force (emf) method. By the use of galvanic cells without liquid junction—that is, containing no regions of contact between two unlike liquid phases—values of pK_A may be determined to a precision of 0.001 to 0.002. Cells

with liquid junction are much more commonly employed by biochemists and are generally more convenient for practical purposes, especially with the complex mixtures with which the biochemist frequently deals. The theory of cells without liquid junctions will be discussed first, however, since it can be discussed more simply and rigorously.

Consider a cell made up of two half-cells connected by a solution of hydrochloric acid at concentration C_1 . One half-cell is a hydrogen electrode. It is made up of a solid core of a noble metal such as gold or platinum, coated with a layer of platinum black; the latter readily absorbs hydrogen, which is bubbled through the solution surrounding the electrode at atmospheric pressure (more strictly the total pressure is equal to the sum of the vapor pressure of the water in the solution and the partial pressure of the hydrogen, and it is this sum which is equal to the atmospheric pressure). The hydrogen in the electrode can react with hydrogen ions in the solution according to the equation



The symbol (g) denotes that the hydrogen is in the gaseous state. The symbol e^- denotes an electron which passes into the electrode and flows into the metallic connections which complete the circuit. The half-cell may be denoted for brevity by the symbols H_2 (1 atm), H^+ , or simply H_2 , H^+ . The other half-cell is made up of a silver electrode, with silver chloride deposited on it in close association with the silver surface. It can react with chloride ions in the solution according to the equation



The symbol (s) in parentheses is used to denote the fact that the silver chloride and silver are in the pure solid state. This half-cell may be denoted for brevity by the symbols Cl^- , AgCl , Ag . The solution between the electrodes must therefore contain both hydrogen (oxonium) and chloride ions. In the simplest case it is a solution of pure hydrochloric acid. The metallic ends of the electrodes are connected by metallic conductors with a galvanometer, and with a potentiometer which can be adjusted to balance the electromotive force of the cell, so that only a minute amount of current flows.⁶ A small adjustment in the potentiometer system can cause current to flow in either direction through the cell, so that the conditions for a truly reversible system are approached very closely indeed—so closely that only a negligible error is involved in the assump-

⁶ The design and use of the potentiometer are considered in detail by W. M. Clark (1928).

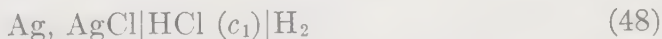
tion that the operation of the cell is strictly reversible in the thermodynamic sense. The entire cell may be written



A comma denotes the junction between two solid phases (H_2 dissolved in platinum black being considered formally as if it were a solid phase), and a vertical line denotes the junction between a solid and a liquid phase. It is assumed in writing such formulas as (46) that the silver and platinum-hydrogen electrodes are suitably connected through an external circuit with a potentiometer. If c_1 is not very large, the electromotive force of cell (46) is such that positive electricity flows through the cell from left to right. This involves a reaction which may be regarded as the sum of reactions (44) and (45):



If $c_1 = 0.1 M$, the electromotive force, E , is $+0.3524$ volt at 25° under the reversible conditions specified above. The positive value for E denotes the fact that the spontaneous flow of positive electric current *through the cell* as written in (46) above is from left to right.⁷ If we wrote the cell as



then by the same convention we should write $E = -0.3524$ volt. Reaction (47) cannot be merely assumed; it must be proved by experimental measurement that, when current flows through cell (46), reaction (47) actually occurs, and that the passage of one faraday, F ($= 96,500$ coulombs), of electricity is accompanied by the formation of one mole of hydrochloric acid; or that passage of current in the reverse direction, due to an applied external electromotive force, leads to an equivalent disappearance of hydrochloric acid, with the formation of silver chloride and hydrogen. This proof has been given by extensive experimentation, in the case of cell (46).

When a cell such as (46) operates reversibly, the maximum electrical work done by the passage of N faraday equivalents (F) of charge, under the influence of the electromotive force, E , is

$$W = -NFE \quad (49)$$

In cell (46), of course, $N = 1$. The gaseous hydrogen at 1 atmosphere, the solid silver chloride, and the metallic silver in equation (47) are all in their standard states, and their chemical potentials are independent of

⁷ The flow of positive current in the external circuit through the wires is of course in the opposite direction, from the hydrogen to the silver-silver chloride electrode.

the concentration of electrolyte in the solution. Hence the electromotive force, at constant pressure and temperature, is a function only of the concentration of hydrochloric acid. The relation may be seen most clearly by considering a double cell containing two compartments with the acid at two different concentrations, c_1 and c_2 :



If $c_1 > c_2$, the spontaneous flow of current will be such that H^+ and Cl^- ions disappear from the left-hand compartment at concentration c_1 and appear in the right-hand compartment at concentration c_2 , with reactions (44) and (45) proceeding from right to left. If the chemical potential of HCl is $(\mu_{\text{HCl}})_1$ at c_1 and $(\mu_{\text{HCl}})_2$ at c_2 , then the free energy change, $-\Delta F$, per mole of H^+ and Cl^- ions disappearing on one side and appearing on the other is equal to $(\mu_{\text{HCl}})_2 - (\mu_{\text{HCl}})_1$. Since the operation of the cell is reversible, this is equal to the work done in the passage of the current, by equation (49). Moreover, we can write for the activity of HCl , by the fundamental relations relating activity and chemical potential (Chapter 4, p. 182),

$$(\mu_{\text{HCl}})_1 - \mu_{\text{HCl}}^\circ = RT \ln (a_{\text{HCl}})_1 \quad (51)$$

and similarly for $(\mu_{\text{HCl}})_2$. The standard state, for which $\mu_{\text{HCl}} = \mu_{\text{HCl}}^\circ$, and $a_{\text{HCl}} = 1$, is fixed by convention. The most convenient convention is to define unit activity of HCl so that a_{HCl} approaches (HCl) as the concentration approaches zero. Combining these relations, from (49), (50), and (51) we obtain

$$\Delta F = -NFE = (\mu_{\text{HCl}})_1 - (\mu_{\text{HCl}})_2 = RT \ln \frac{(a_{\text{HCl}})_1}{(a_{\text{HCl}})_2} \quad (52)$$

The activity coefficient of HCl in dilute aqueous solution varies with the concentration in accordance with the equations of Debye and Hückel (Chapter 5), if it is assumed that the molar concentration of H^+ and Cl^- ions is equal to the molar concentration of added HCl . Therefore, since HCl by this criterion appears to be completely ionized in such solutions, it is natural to represent the chemical potential of HCl as the sum of the chemical potentials of these ions:

$$\mu_{\text{HCl}} = \mu_{\text{H}^+} + \mu_{\text{Cl}^-} \quad (53a)$$

$$a_{\text{HCl}} = a_{\text{H}^+}a_{\text{Cl}^-} = (\text{H}^+)(\text{Cl}^-)\gamma_{\text{H}}\gamma_{\text{Cl}} = (\text{H}^+)(\text{Cl}^-)\gamma_{\pm}^2 \quad (53b)$$

Here the quantities in parentheses denote molar concentrations, γ_{H} and γ_{Cl} denote the activity coefficients of cation and anion, respectively, and $\gamma_{\pm} = (\gamma_{\text{H}}\gamma_{\text{Cl}})^{1/2}$ is the mean ionic activity coefficient of the H^+ and

Cl^- ions. In these terms we may write for the emf of cell (50), taking $N = 1$ in (52),

$$E = \frac{RT}{F} \ln \frac{(\text{H}^+)_1(\text{Cl}^-)_1(\gamma_{\pm}^2)_1}{(\text{H}^+)_2(\text{Cl}^-)_2(\gamma_{\pm}^2)_2} = 0.0591 \log \frac{(\text{H}^+)_1(\text{Cl}^-)_1(\gamma_{\pm}^2)_1}{(\text{H}^+)_2(\text{Cl}^-)_2(\gamma_{\pm}^2)_2} \text{ at } 25^\circ \quad (54)$$

The numerical coefficient $0.0591 = 2.3026RT/F$ is obtained from $R = 8.3144$ joules deg^{-1} mole^{-1} ; $F = 96,493$ coulombs equiv^{-1} , and $T = 298.16^\circ$ K at 25° . The corresponding factor at 30° is 0.06015. We note that, for pure hydrochloric acid, $(\text{H}^+)_1 = (\text{Cl}^-)_1 = c_1$, and correspondingly for c_2 . Hence (54) may be rewritten as

$$E = 0.05916 \log \frac{c_1^2(\gamma_{\pm}^2)_1}{c_2^2(\gamma_{\pm}^2)_2} = 0.1183 \log \frac{c_1(\gamma_{\pm})_1}{c_2(\gamma_{\pm})_2} \quad (55)$$

Thus, if $c_1(\gamma_{\pm})_1$ were ten times as great as $c_2(\gamma_{\pm})_2$, the emf of cell (50) at 25° would be 0.1183 volt, or 118.3 millivolts.

In practice a double cell such as (50) is not so convenient to operate as a single cell like (46). The electromotive force of (46) may be written

$$\begin{aligned} E &= E_0 - \frac{RT}{F} \ln (\text{H}^+)(\text{Cl}^-)\gamma_{\pm}^2 \\ &= E_0 - 0.0591 \log (\text{H}^+)(\text{Cl}^-) - 0.0591 \log \gamma_{\pm}^2 \\ &= E_0 - 0.1183 \log c - 0.1183 \log \gamma_{\pm} \end{aligned} \quad (56)$$

Here E_0 , the standard potential of the cell, is the potential when the hydrochloric acid is in its standard state of unit activity. It may be determined by measuring E at a series of values of $\log c$, down to quite dilute solutions, and plotting $E + 0.1183 \log c$ against $c^{1/2}$. Since $-\log \gamma_{\pm}$ is a linear function of $c^{1/2}$ at low ionic strength—see Chapter 5, equation (76)—the resulting plot is a straight line at very low ionic strength; its extrapolation to $\sqrt{c} = 0$ gives E_0 . Refinements of this procedure, which permit a very accurate extrapolation to obtain E_0 with high precision, are discussed in the monographs of Harned and Owen (1950) and of Robinson and Stokes (1955). The value of E_0 for cell (46) at 25° is $+0.2224$ volt. It is obvious that if we measure E for cell (46), first with HCl at concentration c_1 and then at concentration c_2 , the difference, ΔE , between the two measurements will be the same as the value of E for cell (50).

Consider now what happens in cell (50) when a mixture of electrolytes is employed instead of HCl alone. Suppose that in each compartment we place a mixture of HCl and KCl —say $0.010 M$ HCl and $0.190 M$ KCl in the left-hand compartment, and $0.001 M$ HCl and $0.199 M$ KCl in the right-hand compartment. Reaction (47) at the electrodes is the same as

before, and equation (54) is still applicable. The ionic strength in both solutions is the same (0.200), however, and the potassium and chloride ions are the ions chiefly present in both solutions. Hence, in the light of our discussion of the Debye-Hückel theory in Chapter 5, we should expect that γ_{\pm} for HCl would be nearly the same in both solutions. Likewise the chloride ion concentration is the same in both. Thus the terms in γ_{\pm} and (Cl^-) should cancel in the numerator and denominator of (54), and the emf of the cell should be given simply by

$$E = 0.0591 \log \frac{(\text{H}^+)_1}{(\text{H}^+)_2} \quad (57)$$

Under these special conditions, therefore, the emf of the cell is directly proportional to the logarithm of the ratio of hydrogen ion *concentrations* in the two compartments of cell (50).

Standard Potentials of Half-Cells, Choice of Conventions

It is a matter of great practical convenience to resolve the standard potential of a cell such as (46) into the sum of the standard potentials of two half-cells:

- (a) $\text{H}_2 (g, 1 \text{ atm}) | \text{H}^+, \text{ at unit activity}; E = E_0(\text{H}_2, \text{H}^+)$
 (b) $\text{Cl}^- | \text{AgCl} (s), \text{Ag} (s), \text{ with } \text{Cl}^- \text{ ion at unit activity};$
 $E = E_0 (\text{Cl}^-, \text{AgCl}, \text{Ag})$

and we may write for the potentials at other concentrations

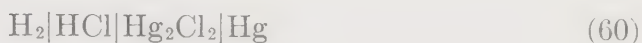
$$\begin{aligned} E(\text{H}_2, \text{H}^+) &= E_0(\text{H}_2, \text{H}^+) - \frac{RT}{F} \ln a_{\text{H}} \\ &= E_0(\text{H}_2, \text{H}^+) - \frac{RT}{F} \ln (\text{H}^+) - \frac{RT}{F} \ln \gamma_{\text{H}^+} \end{aligned} \quad (58)$$

$$\begin{aligned} E(\text{Cl}^-, \text{AgCl}, \text{Ag}) &= E_0(\text{Cl}^-, \text{AgCl}, \text{Ag}) - \frac{RT}{F} \ln a_{\text{Cl}} \\ &= E_0(\text{Cl}^-, \text{AgCl}, \text{Ag}) - \frac{RT}{F} \ln (\text{Cl}^-) - \frac{RT}{F} \ln \gamma_{\text{Cl}^-} \end{aligned} \quad (59)$$

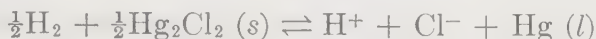
Actually, of course, it is impossible to measure the individual potential of a single half-cell; what is determined experimentally is always the sum of two such half-cell potentials. It is a matter of arbitrary convention to fix the value of E_0 for some standard half-cell. Once this choice is made, the value of E_0 for any other half-cell used in combination with it is fixed by the requirement that the sum of the E_0 values for the two half-cells must be equal to the E_0 value experimentally determined for the com-

bination. The convention universally adopted is that $E_0(\text{H}_2, \text{H}^+)$ is *defined* as having a value of 0 at all temperatures. The value of E_0 for the half-cell Cl^- (unit activity), $\text{AgCl} (s)$, $\text{Ag} (s)$ is then equal to $+0.2224$ volt at 25° . If the standard potentials of two half-cells are known, then the standard potential of the cell made up of a combination of these two half-cells is immediately given, even though this particular cell may not yet have been studied experimentally.

An important electrode which, like the silver-silver chloride electrode, reacts reversibly with chloride ions in solution, is the calomel-mercury electrode: $\text{Cl}^- | \text{Hg}_2\text{Cl}_2 (s) | \text{Hg} (l)$. Very careful measurements have been made on this cell:



for which the total cell reaction may be written



The standard potential for this cell has been determined with high precision as $+0.26796$ volt at 25° , and this is of course by convention equal to the standard potential of the calomel-mercury half-cell at this temperature.

The value 0.26796 volt is given with reference to HCl concentrations expressed as moles per kilogram of solvent (the molality scale), not as moles per liter of solution (the molarity scale). For dilute aqueous solutions, the difference between the two scales is small but not negligible; conversion from one scale to the other can be made if the density of pure water, and of the solution under study, is known at the given temperature. If c is the concentration of solute in moles per liter, and m the molality, then $\lim_{c \rightarrow 0} (c/m) = 1/\bar{v}$, where \bar{v} is the specific volume of water at the given temperature;

\bar{v} is 1.00013 at 0° , 1.00293 at 25° , and 1.00706 at 38° . Thus the limiting value of $\log (c/m)$ at 25° is only 0.0012 , and at 38° only 0.0030 . For such purposes as the calculation of limiting pK_A values, extrapolated to infinite dilution, these differences are almost always negligible in work on biochemical systems. At higher concentrations, of course, the difference between the c scale and the m scale of concentrations becomes increasingly important. In solvents other than water, the difference may obviously be very large, even at infinite dilution.

Individual ionic activity coefficients, such as γ_{H} and γ_{Cl} in (58) and (59), cannot be separately determined by measurements on cells without liquid junction, or indeed by any other method. Certain products of such coefficients, such as $\gamma_{\text{H}}\gamma_{\text{Cl}}$, can be determined experimentally, often with high precision. In the interpretation of some measurements, however—especially those on cells with liquid junctions, which are discussed later—it is often convenient to define such individual ionic activity coefficients

by some suitable convention. This point is discussed further below, in connection with pH measurements on cells with liquid junction.

EVALUATION OF pK_A FROM CELLS WITHOUT LIQUID JUNCTION

Cells such as (46) and (50) have been employed, especially by H. S. Harned and his associates, to determine the pK_A constants of weak acids. One may, for example, set up a cell containing a buffer mixture of an acid, HA , and its salt, Na^+A^- , with an added chloride such as Na^+Cl^- :



Here the hydrogen ions, which react reversibly at the hydrogen electrode, are derived from HA , the chloride from Na^+Cl^- . Equation (56) gives the electromotive force. We may write the ionization constant of the acid as

$$K_A = \frac{\gamma_H \gamma_A (H^+) (A^-)}{\gamma_{HA} (HA)} \quad (62)$$

Combining (56) and (61), and remembering that γ_{\pm}^2 in (56) denotes the activity coefficient product $\gamma_H \gamma_{Cl}$, we can cancel (H^+) from the equation for the emf and separate the terms involving the molar concentrations of Cl^- , HA , and A^- from the terms involving activity coefficients:

$$E - E_0 + 0.0591 \log \frac{(Cl^-)(HA)}{(A^-)} = -0.0591 \log \left(\frac{\gamma_H \gamma_{Cl}}{\gamma_H \gamma_A} \gamma_{HA} K_A \right) \quad (63)$$

Moreover $(Cl^-) = m_3$ in (61), $(A^-) = m_2 + (H^+)$, and

$$(HA) = m_1 - (H^+)$$

For acids with $K_A < 10^{-5}$, it is a good approximation to set $(A^-) = m_2$, and $(HA) = m_1$. Even for somewhat stronger acids we can assume this as a first approximation, and then use it to obtain more accurate values later, as in equations (30) to (33) above. Then (63) becomes

$$E - E_0 + 0.0591 \log \frac{m_3 m_1}{m_2} = -0.0591 \log K_A - 0.0591 \log \left(\frac{\gamma_H \gamma_{Cl}}{\gamma_H \gamma_A} \right) \gamma_{HA} \quad (64)$$

All terms on the left can be experimentally determined, since E , m_1 , m_2 , and m_3 are directly determined, and E_0 is already known from an independent set of measurements, as previously described. In the activity coefficient terms on the right, $-\log \gamma_H \gamma_{Cl}$ is a linear function of the square root of the ionic strength ($\sqrt{\omega}$) at low ionic strengths, and so is $-\log \gamma_H \gamma_A$ (see Chapter 5). Their ratio, however, is not far from unity

at ionic strengths below 0.15 or thereabouts and is approximately linear in ω , as is γ_{HA} . Hence, if cell (61) is studied over a range of ionic strengths, obtained by varying the sodium chloride concentration, and the left-hand side of (64) plotted against the ionic strength, the intercept at $\omega = 0$ gives $-0.0591 \log K_A$, since the last term on the right of (64) must vanish as ω goes to zero. By essentially this method, Harned and Ehlers found for acetic acid $K_A = 1.754 \times 10^{-5}$ at 25° , and by the same procedure determined K_A over a whole range of temperatures from 0° to 60° . The same method has been applied to a number of carboxylic acids and amino acids to give very accurate K_A values over a range of temperatures. The details of procedure are given by Harned and Owen, and by Robinson and Stokes; we have merely outlined the principle of the method here.⁸ Some numerical values of pK_A will be found on pp. 452-453.

Cells Containing Liquid Junctions

The galvanic cells employed by biochemists generally, and sometimes by physical chemists, in measurements concerned with acid-base equilibria involve a junction between two different liquid phases.

The reader may raise the question: Why make use of cells with liquid junction, if measurements on cells without liquid junction are admittedly more accurate and reproducible? For the biochemist, the answer in many cases is clear. Substances such as proteins and nucleic acids, and many simpler substances of biochemical interest, form complexes with the ions of heavy metals. Silver-silver chloride or mercury-calomel electrodes, if brought into direct contact with such solutions in a cell without liquid junction, would interact with such compounds. The electrode reactions would no longer be precisely defined, and the substances in the solution would be changed by their reactions with the electrode materials. These complications are avoided by the use of a liquid junction of potassium chloride solution, interposed between the solution under study and the silver or mercury electrode. It is also generally true that measurements on cells without liquid junction are more cumbersome, and a considerably longer time is required to attain equilibrium, than in cells with liquid junction of the type described below. Often a rapid approximate measurement, to indicate the state of acid-base equilibria in a given system, is much more useful—for instance in following the course of a biochemical preparation—than a more accurate measurement by more elaborate procedures.⁹

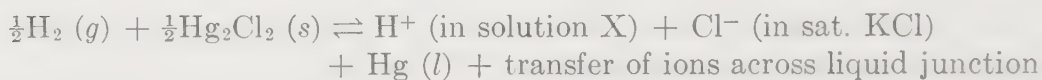
⁸ The cells employed, in the hands of skillful workers, can give electromotive forces which are reproducible within a few hundredths of a millivolt, and pK_A values can be determined within ± 0.001 . Some of the values so obtained are listed in Table III.

⁹ Even in a cell without liquid junction, the liquid phase is not completely uniform throughout the region between the two electrodes. For instance, in the cell $\text{H}_2|\text{HCl}|\text{AgCl}, \text{Ag}$ the portions of the liquid which adjoin the two electrodes are kept more or less separate from the main body of the solution. It is necessary to keep the gaseous hydrogen from coming into contact with the silver-silver chloride electrode, lest its reducing action should disturb the functioning of the cell. On the other hand, the liquid close to the silver-silver chloride electrode must contain an appreciable amount

The commonest example of a cell with liquid junction is a cell consisting of two half-cells—one a metal-hydrogen electrode of the type already described, in contact with the solution under study, which we may call solution X; and the other, of an electrode of metallic mercury and calomel (Hg_2Cl_2). The latter, like the silver-silver chloride electrode, reacts reversibly with chloride ions, and must therefore be in contact with a solution containing chloride ions. The usual choice is a potassium chloride solution; and in pH measurements as usually carried out today it is made up of saturated potassium chloride. The cell therefore contains two different liquids, in contact with the two electrodes; and a junction between the two unlike liquids must therefore be formed, in order to complete the circuit. For brevity, such a cell may be denoted by the symbols



where the double vertical line denotes the liquid junction. The cell reaction is



Since $E_0 = 0$ by definition for the hydrogen electrode, the potential of the electrode on the left is, from (58), equal to $-(RT/F) \ln a_{\text{H}}$. The potential of the half-cell on the right may be written as E^* , which is the value of E in (59) when a_{Cl} is set equal to the activity of chloride ion in saturated potassium chloride. The liquid junction potential may be designated by E_j . Then for cell (65)

$$E = -\frac{RT}{F} \ln a_{\text{H}} + E^* + E_j \quad (66)$$

or

$$-\log a_{\text{H}} = \frac{E - E^* - E_j}{(2.3026RT/F)} = -\log [(\text{H}^+)\gamma_{\text{H}}] \quad (66a)$$

The value of E^* is a constant, independent of the composition of solution X. If the same were true of E_j , we could then determine the pH of solution X, with reference to some suitably chosen standard solution,

of silver ion, the amount being determined by the chloride ion concentration and the solubility product of silver chloride. The silver ion concentration in the body of the liquid and at the hydrogen electrode is, however, essentially zero. The concentration of silver ion at *any* point in the liquid is so small, however, that for most purposes it can be neglected as far as its effect on the activities of other components is concerned. It will be seen from these considerations that the composition of the liquid is not homogeneous in the strictest sense of the word, even in a cell without liquid junction.

the pH of which is fixed by definition. If pH_s is the pH of the standard solution, pH_x the pH of solution X, and E_s and E_x the emf values of these two solutions when measured in cell (65), then from (66a), setting $pH = -\log a_H$,

$$pH_x = pH_s + \frac{E_x - E_s}{(2.3026RT/F)} = pH_s + \frac{E_x - E_s}{0.0591} \text{ at } 25^\circ \quad (67)$$

This is in fact the procedure generally adopted for the practical determination of pH . To justify and clarify the procedure, however, we must consider several important questions: (1) How accurate is the assumption that E_j is the same for the standard solution and for solution X? (2) How do pH measurements vary with ionic strength and other variables, in acids of different charge types? (3) What criteria are to guide the choice of the standard solution, or solutions?

THE LIQUID JUNCTION POTENTIAL

This potential, E_j , is determined by the work done in transporting all the ions across the junction when current flows. Since solvent molecules may also be transported when current flows, the free energy change occurring when solvent is transported across the boundary must be included also in the work for the whole process. If the transference number of an ion (Chapter 7) is t_i , and its valence Z_i , then t_i/Z_i moles of this ion move across the boundary in the direction of positive current when one equivalent of electricity flows through. Let the molar increment in chemical potential of an ion, when it is moved through an infinitesimal region in the solution, be $d\mu_i$. Then the work done in moving n_i moles of such ions is $n_i d\mu_i = RTn_i d \ln a_i$, where a_i is the activity of the ion. (Here we apply to each individual ion the relation between activity and chemical potential derived for a component of a system in Chapter 4, p. 184; that is, in differential form, the relation is $d\mu_i = RT d \ln a_i$.) Hence the work done in transporting t_i/Z_i moles of the ion between solution I (solution X) and solution II (saturated KCl), through the entire boundary region, is given by integrating this expression:

$$W_i = RT \int_I^{II} d \frac{t_i}{Z_i} \ln a_i \quad (68)$$

If t_s moles of solvent are also transported per equivalent of electricity, then there must be an additional term equal to $RT \int_I^{II} t_s d \ln a_s$. The total work, W , is the sum of all such terms for ions and solvent. The

liquid junction potential, E_j , is, from (49), equal to $-W/\mathbf{F}$, since we have taken $N = 1$ for the transport process. Then

$$E_j = -\frac{RT}{\mathbf{F}} \int_{\text{I}}^{\text{II}} \sum d \frac{t_i}{Z_i} \ln a_i = -\frac{RT}{\mathbf{F}} \int_{\text{I}}^{\text{II}} \sum d \frac{t_i}{Z_i} \ln (\gamma_i c_i) \quad (69)$$

The term for the solvent is understood to be included in (69). Since t_i and a_i for each ion will in general vary from point to point across the boundary, as the composition of the solution changes from solution I at one end to solution II at the other, the evaluation of the terms in the integral of (69) is obviously no simple matter. Indeed a fundamental paradox emerges. In order to solve (69) for E_j , we must know the activity coefficient, γ_i , for each ion, including the hydrogen ion. But in order to determine the activity coefficient of the hydrogen ion, by (66) or (66a), we must know the liquid junction potential, E_j . This situation confronts us with an apparently inescapable dilemma, and we must conclude that there is no experimental procedure which can lead to an unambiguous determination of individual ion activities by any thermodynamic method.

Important conclusions can be drawn, however, from equation (69), which are very helpful as a guide to experimental procedure in the operation of cells such as (65). The value of t_i for each ion is proportional to the concentration of the ion at the junction, and also to the electrical mobility of the ion. If the principal ion constituents of the boundary were only two in number, with $t_i = 0.5$ for each, and with Z_i of course of opposite sign for the two, then each of these two ions would be expected to contribute a nearly equal term—one positive, one negative—to the integral of (69), and the total value of E_j would be nearly zero. These requirements are approximately fulfilled by the use of a saturated solution of potassium chloride. The potassium and the chloride ions both have the electronic configuration of argon, and the somewhat greater size of the chloride ion is essentially compensated by the greater hydration of the potassium ion. Hence $t_+ \cong t_-$ and E_j should be quite low. The ions present on the other side of the liquid junction, however—that is, for cell (65), in solution X—inevitably make some contribution to E_j . If the total concentration of such ions is small, and if they are not of high mobility—which means that the concentration of the highly mobile H^+ and OH^- ions should be very small—then the contribution of the ions of solution X to the liquid junction should be a minor one.

It is not necessary, however, nor indeed possible, that the liquid junction potential should be completely eliminated. It is important, however, that it should be nearly the same for the standard and for the solution under test. This means that the mobilities and numbers of the ions present in the standard and unknown solutions should preferably not be

very different. It is also desirable that the pH values of the two should not be too far apart—preferably not more than 1 or 2 pH units. To fulfill this requirement, of course, several standards of different pH values must be available.

Liquid junction potentials are likely to be moderately large in solutions containing appreciable amounts of hydrogen or hydroxyl ions—we may say, roughly, at pH values below 3 or above 11. If many H^+ ions are present at a junction of the type $HCl (0.1 M) | KCl (sat.)$, the rapidly moving H^+ ions outdistance the Cl^- ions, in effect setting up an electric double layer, the positive side of which is in the saturated KCl solution. The magnitude of E_j has been estimated as close to 3 millivolts for such a boundary from (69), making plausible assumptions about the composition of the boundary layers between the two solutions, and about the activity coefficients of the individual ions. At the junction $NaOH (0.1 M) | KCl (sat.)$ the OH^- ions move faster than the Na^+ ions, and the charge distribution in the double layer is opposite to that shown above for HCl . In this case E_j has been estimated as of the order of -2 millivolts. Such estimates are always uncertain, but those given are almost certainly of the right order of magnitude.

The subject of liquid junction potentials has been examined extensively in the literature, on which we have not touched here. The reader who wishes to inquire further into the subject will find detailed discussion, and extensive references to the literature, in the following books cited at the end of this chapter: Bates, Chapters 3 and 7; Harned and Owen, Chapter 10. A discussion by Scatchard, in Cohn and Edsall's "Proteins, Amino Acids and Peptides," pp. 39–45, is also recommended.

VARIATION OF pH WITH IONIC STRENGTH FOR ACID-BASE PAIRS OF DIFFERENT CHARGE TYPES

Accurate values of pK_A obtained from cells without liquid junction, or from conductivity or indicator measurements, carried out with the necessary corrections and precautions, have furnished a precise basis for the evaluation of acid-base equilibria in many systems. We now wish to correlate these data with the usual measurements made on cells with liquid junctions, so as to establish a pH scale that will be consistent with them. Necessarily, pH measurements are carried out on systems at finite ionic strength, and the extrapolation to zero ionic strength is attended with some uncertainty in systems containing liquid junctions. It is convenient to define an "apparent K value," denoted by K' , for the equilibrium $A \rightleftharpoons B + H^+$, by the relation

$$K_A = \frac{a_H a_B}{a_A} = a_H \frac{(B)\gamma_B}{(A)\gamma_A} = K'_A \frac{\gamma_B}{\gamma_A} \quad (70)$$

Hence, defining pH as $-\log a_H$, we obtain

$$pH = pK'_A + \log \frac{(B)}{(A)} = pK_A + \log \frac{(B)}{(A)} + \log \frac{\gamma_B}{\gamma_A} \quad (71)$$

The value of pK_A is known with high precision from measurements on cells without liquid junction. The concentrations (B) and (A) can also be determined with high precision from the stoichiometry of the system, with corrections if necessary for the effect of reactions such as $A + H_2O \rightleftharpoons B + H_3O^+$ (see equations 30 to 33 above). We must then make some reasonable assumption concerning $\log \gamma_A$ and $\log \gamma_B$. Either A or B must be electrically charged, and both may be.

At moderate ionic strengths we may try to evaluate the activity coefficients of ionic constituents by applying the Debye-Hückel equation, expanded by the inclusion of a term linear in the ionic strength, involving a salting-out constant, K_s (compare Chapter 5, equation (76.5):

$$-\log \gamma_i = \frac{0.5z_i^2 \sqrt{\omega}}{1 + Qa \sqrt{\omega}} - K_s \omega \quad (72)$$

Here a may be taken as the "collision diameter" in angstroms, $Q \cong 0.33$, and the value of Qa lies usually between 1 and 2. If the charge of an acid, A , is Z , then the charge of the conjugate base, B , is $Z - 1$. Moreover, at relatively low ionic strengths, it is often a useful approximation to assume that the salting-out term, which depends on the size and general structure of the ion, but not on its net charge, is the same for the acid and for its conjugate base, so that it cancels in the expression $\log (\gamma_B/\gamma_A)$ of (71). We also assume that the value of a is the same for an acid and for its conjugate base. Then we can write, for an acid of net charge Z and a conjugate base of net charge $Z - 1$,

$$\begin{aligned} pK' - pK &= \log \frac{\gamma_B}{\gamma_A} = \frac{0.5[(Z - 1)^2 - Z^2] \sqrt{\omega}}{1 + Qa \sqrt{\omega}} \\ &= \frac{-0.5(-2Z + 1) \sqrt{\omega}}{1 + Qa \sqrt{\omega}} \end{aligned} \quad (73)$$

Thus, for the acetic acid-acetate system ($Z = 0$) the numerical coefficient of $\sqrt{\omega}$ in the numerator of (73) is -0.5 . For the system primary phosphate-secondary phosphate, involving the reaction:

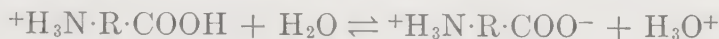


$Z = -1$, and the coefficient of $\sqrt{\omega}$ is -1.5 . For an alkylammonium ion and its conjugate base, $Z = +1$, and the numerical coefficient in (73) is $+0.5$.

If either the acid or its conjugate base is a dipolar ion, its net charge is zero, but its very large dipole moment leads to a salting-in term proportional to ω but of opposite sign to K_s . We may therefore, in the light of the detailed discussion given in Chapter 5, represent the logarithm of the activity coefficient of a dipolar ion by an equation of the form

$$-\log \gamma \text{ (dipolar ion)} = (K_R - K_s)\omega \quad (72a)$$

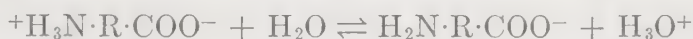
at low ionic strengths. Roughly speaking, K_R , the salting-in term, is large if the dipole moment of the dipolar ion is large, and K_s is large if the molecule contains large nonpolar residues. For an α -amino acid in water, K_R is of the order of 0.32; for a dipeptide, of the order of 0.6. A simple dipolar ion ($^+\text{H}_3\text{N}\cdot\text{R}\cdot\text{COO}^-$) is a base, to which the cation ($^+\text{H}_3\text{N}\cdot\text{R}\cdot\text{COOH}$) is the conjugate acid; it is also an acid to which an anion ($\text{H}_2\text{N}\cdot\text{R}\cdot\text{COO}^-$) is the conjugate base. As in the previous discussion concerning ions, we assume as a first approximation that the salting-out term (K_s) is the same for any acid and for its conjugate base. We may use equation (72) for the cation or anion, and (72a) for the dipolar ion. Then we obtain for pK_1 of an amino acid or peptide, corresponding to the reaction



the equation

$$pK_1' = pK_1 + \frac{0.5 \sqrt{\omega}}{1 + Qa \sqrt{\omega}} - K_R\omega \quad (73a)$$

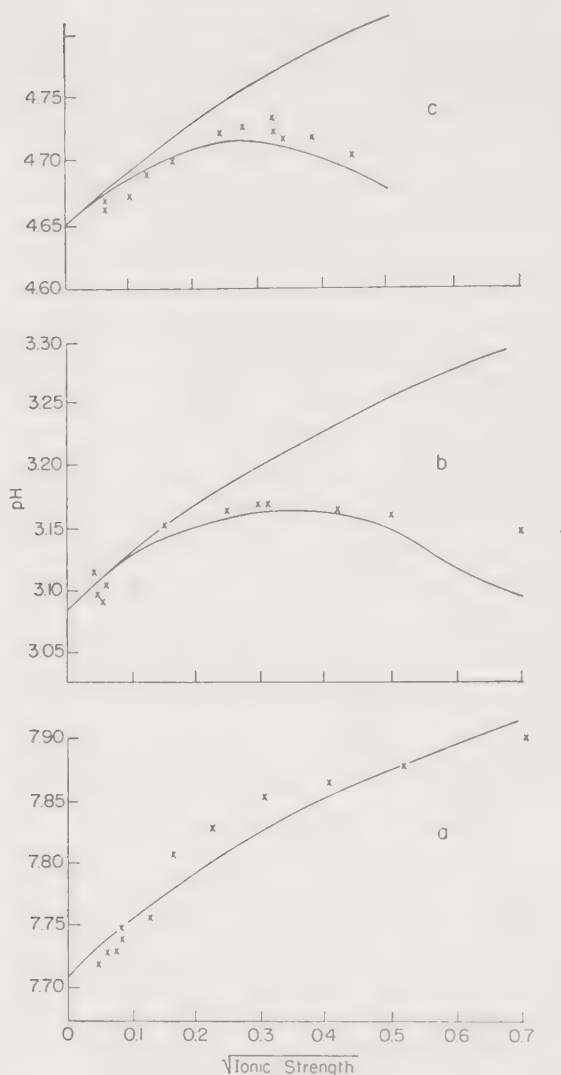
and for pK_2 of an amino acid or peptide, corresponding to the reaction



the equation

$$pK_2' = pK_2 - \frac{0.5 \sqrt{\omega}}{1 + Qa \sqrt{\omega}} + K_R\omega \quad (73b)$$

Equations (73), (73a), and (73b) have been used by Neuberger (1937) for calculating the pK' values for acids of different charge types, including dipolar ions. For simplicity Neuberger set $Qa = 1$ in these equations, and his data, shown in Figs. 5 and 6, show the general agreement between the experimental pH measurements and the expectation from the equations. This kind of agreement, found also in a great variety of other studies on other acid-base systems, indicates the usefulness of the concept of the activity coefficient of an individual ion, and its computation—at low ionic strengths—by means of the Debye-Hückel theory.

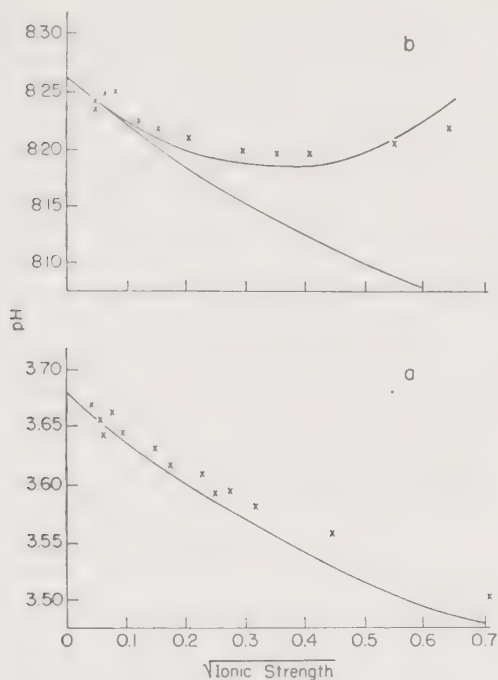


FIGS. 5a, 5b, and 5c. x, Experimental points. a, glycyl-glycine ester; b, glycylglycine pK_1 ; c, ϵ -aminododecanoic acid. The curve in Fig. 5a and the upper curves in Figs. 5b and 5c have been plotted according to equation (73), taking $Z = 0$. The lower curve in Figs. 5b and 5c are plotted according to equation (73a), with K_R equal to 0.4 in Fig. 5b and to 0.55 in Fig. 5c. The value of Qa has been taken as equal to 1 in all cases. (From A. Neuberger, 1937.)

CHOICE OF A STANDARD FOR CALIBRATING pH MEASUREMENTS

Such data as those shown in Figs. 2 and 3 indicate the basis for an approach to the problem of calibrating pH measurements so as to give data which are to be as consistent as possible with the measurements on cells without liquid junction.

In the calibration procedure, electromotive force measurements are made on a cell of the type of (65), and the pH difference between any two of the measured solutions is calculated from the measurements by



FIGS. 6a and 6b. x , Experimental points. a, *N*-acetylglycine; b, glycylglycine pK_2 . The curve in Fig. 6a and the lower curve in Fig. 6b are plotted according to equation (73), taking $Z = +1$. The upper curve in Fig. 6b is plotted according to equation (73b) with K_R equal to 0.4. As in the curves of Fig. 5, $Qa = 1$ in all the curves drawn. (From A. Neuberger, 1937.)

(67). We then choose an absolute value of pH , for some one buffer mixture taken as a standard, so as to get an extrapolated value of pK (equation 73) as consistent as possible with the data on cells without liquid junction. Once an absolute value for one solution is chosen by convention, equation (67) fixes the pH values for other solutions by reference to this standard, since pH_s and the corresponding value of E_s in (67) are now given. It is an experimental fact that consistent results are then obtained by pH measurements on other buffer mixtures, with different acids and their conjugate bases, so that the pK values obtained by extrapolation—using equations (73), (73a), or (73b), according to the nature of the acid—are all consistent with the data from cells without liquid junction, within approximately ± 0.01 pH unit. The decision as to which particular buffer mixture should be chosen as a standard is made largely on the basis of convenience. The components of the standard mixture should be readily available in pure form, and the procedure for making up the buffer should be simple. In practice it is useful to have several working standards covering different parts of the pH range.

Indeed the experimental measurements show, for an uncharged acid such as acetic acid, in buffer mixtures with its conjugate base at various

ionic strengths, that the value of pK' varies with the square root of the ionic strength approximately according to equation (73), and that the limiting slope of the curve for pK' as a function of $\sqrt{\omega}$, at ω values around 0.01 to 0.03, is not far from the value of 0.5 predicted by the Debye-Hückel theory. Similar statements apply approximately to acids of other charge types, as shown by Figs. 5 and 6. In obtaining the best extrapolation, it is legitimate to regard Qa , in equations (73), (73a), and (73b), as an adjustable parameter, rather than fixing it as equal to 1, as in Neuberger's studies. The limiting slope of the curve may deviate somewhat from the theoretical Debye-Hückel value, unlike similar curves obtained from cells without liquid junction, where the agreement is excellent. There are likely to be irregularities at low ionic strengths—below 0.01 or thereabouts—so that if the pH measurements are calibrated to extrapolate to the correct pK value, using equations (71) and (72) or (73) between $\omega = 0.01$ and 0.15, the results may become inconsistent below $\omega = 0.01$. This is probably due to changes in liquid junction potentials. At considerably higher ionic strengths, in buffer solutions—if ω rises much above 0.2 or thereabouts—the liquid junction potential must increase, and the physical meaning of the recorded emf values become increasingly uncertain as ω increases. Measurements can still be made, of course, and recorded as pH values in terms of a suitable reference standard by (67). Such values are often empirically useful. The correlation of pH measurements, however, with accurate pK values on cells without liquid junction is significant chiefly in the range from about $\omega = 0.01$ to $\omega = 0.15$ to 0.20.

The general approach to the calibration of pH measurements, as outlined here, has been due to the work of E. J. Cohn, D. A. MacInnes, D. I. Hitchcock, and others. Actually they approached the problem in a somewhat different way from that outlined here, by attempting to find a suitable value for $E^* + E_j$ in equation (66a), which could be used consistently for a cell such as (65) at a given temperature. Their procedure, however, leads logically to the choice of certain solutions as standards of reference. The values chosen for several standard solutions by MacInnes and by Hitchcock coincide very closely with those obtained in the National Bureau of Standards in Washington by a somewhat different but entirely consistent method of calibration, which will be described below. As one example of the consistency of the results, we may quote the pH values chosen for a standard buffer consisting of 0.1 M acetic acid and 0.1 M sodium acetate at 25°.

D. I. Hitchcock and A. C. Taylor

D. A. MacInnes, D. Belcher, and T. Shedlovsky

National Bureau of Standards

pH 4.645

pH 4.640

pH 4.65

All agree within ± 0.01 in pH , which for most purposes may be taken as the limit of significance for pH measurements, although a reproducibility higher than this may be achieved by a single careful operator. The value of pK for acetic acid at 25° is 4.756; the use of equations (71) and (73), based on a pH value of 4.65 for the buffer indicated above, leads to the choice of $Qa = 1.5$ in (73), or $a = 4.6$ A, approximately. This is a plausible value for this empirical parameter.

The standardization procedure employed at the National Bureau of Standards (NBS) has been described in detail by R. G. Bates (1954). It involves measurements on cells without liquid junction, with hydrogen and silver-silver chloride electrodes, of the type of (46), with a buffer solution containing added chloride ion between the electrodes. From equation (56) and making use of the definition $\gamma_H\gamma_{Cl} = \gamma_{\pm}^2$, we may define a quantity $pwH = -\log [\gamma_H\gamma_{Cl}(H^+)]$ by the relations

$$\begin{aligned} \frac{(E - E_0)\mathbf{F}}{2.3026RT} &= -\log (H^+) - \log (Cl^-) - \log (\gamma_H\gamma_{Cl}) \\ &= pwH - \log (Cl^-) \quad (74) \end{aligned}$$

The added chloride ion concentration (Cl^-) is a known quantity, E is directly measured, and E_0 is determined by the methods previously outlined. Therefore pwH can be determined with high precision. It is evaluated in a series of solutions of the buffer under study, at a series of chloride ion concentrations, and pwH is plotted against (Cl^-) or against (Cl^-)/ ω . The extrapolated value of pwH at (Cl^-) = 0 is denoted by pwH° . If the buffer concentration is varied along with (Cl^-), so as to keep the total ionic strength constant, the terms in $\log (\gamma_H\gamma_A)$ which involve $\sqrt{\omega}$ should stay nearly constant, and pwH should be very nearly a linear function of ω . Experiment shows that this is indeed the case. In fact, a linear plot, though with a somewhat steeper slope, is also obtained if the ionic strength of the buffer is held constant in a series of measurements, and only the chloride ion concentration is varied. The same extrapolated value of pwH° , within ± 0.001 , is obtained in either case. Thus for a standard phosphate buffer—0.025 M KH_2PO_4 and 0.025 M Na_2HPO_4 —a value of $pwH^\circ = 6.972$ at 25° was obtained by either procedure.

The final step in translating such data into pH values inevitably involves the introduction of the activity coefficient of an individual ion—in this case the chloride ion—by some reasonable assumption, to give the pH of the standard buffer solution, pH_s :

$$pH_s = pwH^\circ + \log \gamma_{Cl}^\circ \quad (75)$$

where γ_{Cl}° is the activity coefficient of chloride ion, extrapolated to zero chloride concentration, in the buffer of ionic strength ω . It may be com-

puted on a number of different assumptions, several of which are discussed in detail by Bates. One plausible assumption is to use equation (72) for $\log \gamma_{\text{Cl}}^\circ$, choosing a reasonable value of Qa , such as 1.5. Fortunately at ionic strengths below 0.1 all the assumptions examined by Bates led to the same value of pH (6.86) at 25° for the phosphate buffer

TABLE II
VALUES OF pH FOR CERTAIN STANDARD SOLUTIONS FROM 0° TO 95°
INCLUDING VALUES OF pK_w FROM 0° TO 60°

Temperature, ($^\circ\text{C}$)	0.05 M K-tetroxalate	KH tartrate (satd. at 25°)	0.05 M KH phthalate	0.025 M KH ₂ PO ₄ , 0.025 M Na ₂ HPO ₄		0.01 M borax	pK_w
0	1.67	—	4.01	6.98		9.46	14.943
5	1.67	—	4.01	6.95		9.39	14.734
10	1.67	—	4.00	6.92		9.33	14.535
15	1.67	—	4.00	6.90		9.27	14.346
20	1.68	—	4.00	6.88		9.22	14.167
25	1.68	3.56	4.01	6.86		9.18	13.996
30	1.69	3.55	4.01	6.85		9.14	13.833
35	1.69	3.55	4.02	6.84		9.10	13.680
40	1.70	3.54	4.03	6.84		9.07	13.535
45	1.70	3.55	4.04	6.83		9.04	13.396
50	1.71	3.55	4.06	6.83		9.01	13.262
55	1.72	3.56	4.08	6.84		8.99	13.137
60	1.73	3.57	4.10	6.84		8.96	13.017
70	—	3.59	4.12	6.85		8.92	
80	—	3.61	4.16	6.86		8.88	
90	—	3.64	4.20	6.86		8.85	
95	—	3.65	4.22	6.87		8.83	

From R. G. Bates, "Electrometric pH Determinations," p. 74, Table 4, John Wiley & Sons, New York (1954); pK_w values from Harned and Owen (1950). Recently, R. G. Bates, V. E. Bower, and E. R. Smith, *J. Research Natl. Bur. Standards* **56**, 305 (1956) have recommended saturated calcium hydroxide solution as a pH standard for highly alkaline solutions. The solution at 25° is about 0.0203 M , with a pH of 12.45. Revised values of pH for a series of buffer mixtures, originally developed by W. M. Clark and H. A. Lubs, have been reported by V. E. Bower and R. E. Bates, *ibid*, **55**, 197 (1955).

specified above, within ± 0.01 . Use of bromide ions in the solution, with silver-silver bromide electrodes, or of iodide ions with silver-silver iodide electrodes, in conjunction with the same buffer solutions, led to the same final pH within the same limit of error.

Analogous studies with various buffer systems, at various temperatures, yielded a series of pH values which are listed in Table II. The

standard buffers were chosen to cover a wide pH range, the substances composing them being readily available and easily purified. Where comparisons have been made, the values in Table II agree within ± 0.01 with values previously derived for the same solutions by MacInnes or Hitchcock, on the basis of the calibration described earlier. This is true, except for a few quite acid solutions, of pH 2 or below, where the deviation is larger, of the order of 0.02. Measurements of pH should always be referred to suitable standards, such as those of Table II, and the accurate measurement of the standard solution is obviously just as important as that on the unknown under study, if the pH value of the latter is to be properly defined.

The Glass Electrode

Most pH measurements today are not carried out on the hydrogen electrode (cell 65) but on electrode systems which contain a membrane made of a suitable type of glass which functions, over a wide pH range, as if it were a hydrogen electrode. The solution to be studied is placed on one side of the glass membrane; it is connected by a bridge of saturated potassium chloride to a calomel-mercury electrode in the manner already described. The other side of the glass membrane is in contact with a solution of fixed pH and chloride ion content, usually dilute hydrochloric acid, which is in contact with an electrode reversible to chloride ion. The assembly may be denoted in outline by



Over the pH range in which it functions normally, emf measurements on a glass electrode, first with a standard solution of known pH and then on the unknown under study, are related by equation (67) just as they are with the hydrogen electrode. Measurements on the hydrogen electrode remain as the ultimate basis of calibration, of course, but the glass electrode, after a long and difficult development, has evolved into an instrument on which measurements can be made reliably and with ease. Substances which are reduced by gaseous hydrogen cannot be studied on the hydrogen electrode, but they can be handled without difficulty on the glass electrode. On systems like hemoglobin solutions, which interact with gases such as oxygen and carbon dioxide, pH measurements can be carried out readily at defined gas pressures with the glass electrode, but only with great difficulty or not at all on the hydrogen electrode.

The high electrical resistance of the glass membrane was a source of considerable difficulty in earlier work; the glass was often blown into rather large, very thin bulbs of relatively low resistance. With the development of adequate electronic amplifier systems for measuring voltage

in such systems, these difficulties have receded, and more rugged electrodes of higher resistance are now generally used. The choice of the glass is of great importance. It should have relatively low resistance, it should absorb considerable amounts of water in the interstices of the silica network to permit ready exchange of protons between the solution and the interior of the glass, and it should of course give emf values which vary with pH , when used with cell (76), in the same manner as the emf of a hydrogen electrode, cell (65). No one glass meets all these requirements at all pH values between 0 and 14. D. A. MacInnes and M. Dole chose a glass containing 72.2 mole % SiO_2 , 6.4 mole % CaO , and 21.4 mole % Na_2O ; this is now manufactured by the Corning Glass Works and known as Corning 015 glass. It is excellent for work in the pH range between approximately 1 and 8.5; but particularly in alkaline solutions above 8.5 to 9 it shows errors in solutions containing sodium ions. These errors increase as the pH increases and may be as great as 1 pH unit or more between pH 12 and 13 in the presence of molar sodium ion concentration. The error is such that the measured pH is lower than the true pH established on a hydrogen electrode. Evidently the glass electrode functions to some extent as a sodium electrode in the presence of large concentrations of Na^+ and very small concentrations of H_3O^+ . Glass electrode errors in solutions containing potassium ions but no sodium are very much less.

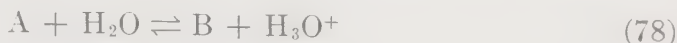
The development of other types of electrode glass containing Li_2O instead of Na_2O has greatly reduced the alkaline errors, and true pH values can be obtained in solutions as alkaline as pH 12.5, even with considerable sodium ion present. Lithium ions are apparently more strongly bonded than sodium ions within the silica network and do not exchange so readily with cations in the solution. It has been reported that the addition of some cesium or rubidium oxide to the lithium oxide in preparing the glass is effective in further reducing the sodium error. The use of lanthanum oxide instead of calcium oxide in the glass has also been recommended.

Determination of pK_A Values and Related Thermodynamic Functions in Relation to Structure

Data of high accuracy are now available, obtained chiefly from measurements on cells without liquid junction, for a number of acids, giving pK_A values over a range of temperature, usually between 0° and 50° or 60° . From these data it has been possible to evaluate not only the standard free energy change, ΔF° ,

$$\Delta F^\circ = -RT \ln K_A = 2.3026RTpK_A \quad (77)$$

but also the molal changes in heat content (ΔH°), entropy (ΔS°), and heat capacity at constant pressure (ΔC_p°) which occur when the reaction



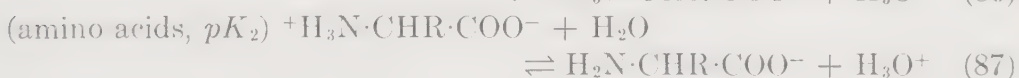
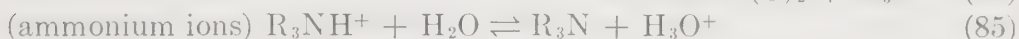
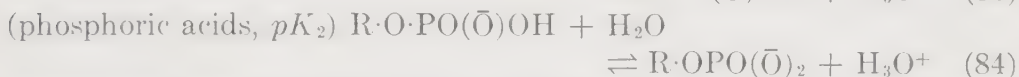
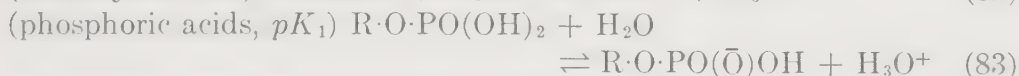
occurs under standard conditions, A and B denoting an acid and its conjugate base. Pure water is taken as the standard state of unit activity for H_2O ; for A, B, and H_3O^+ the standard state of unit activity is defined by the convention that the limit a/m approaches unity as m approaches zero. The fundamental formulas are developed in Chapter 4 but are repeated here for convenience. (It is understood that all the partial derivatives below are taken at constant pressure.)

$$\Delta H^\circ = RT^2 \frac{\partial \ln K_A}{\partial T} = -R \frac{\partial \ln K_A}{\partial (1/T)} = 2.3026R \frac{\partial pK_A}{\partial (1/T)} \quad (79)$$

$$\Delta S^\circ = - \frac{\partial \Delta F^\circ}{\partial T} = \frac{\Delta H^\circ - \Delta F^\circ}{T} \quad (80)$$

$$\Delta C_p^\circ = \frac{\partial \Delta H^\circ}{\partial T} \quad (81)$$

A selected series of typical values for substances of general chemical and biochemical interest is given in Table III. The reactions involved are:



Here R may denote a hydrogen atom or any one of various organic groups. For aspartic acid, pK_1 and pK_2 refer to the removal of protons from the two carboxyl groups, and pK_3 to the removal of a proton from the $-\text{NH}_3^+$ group. For carbonic acid pK_1 refers to an over-all reaction, involving two steps:

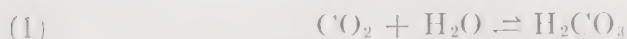


TABLE III
 pK_A VALUES AND RELATED THERMODYNAMIC FUNCTIONS FOR CERTAIN
 CARBOXYLIC ACIDS, PHOSPHORIC ACIDS, AMMONIUM IONS, AMINO
 ACIDS, AND CARBONIC ACID
 (All data for 25°)

Substance	pK_A	ΔF°	ΔH°	ΔS°	ΔC_p°	Ref.
Water (K_w)	13.997	19089	13519	-18.7	-47	1
Carboxylic acids						
Formic acid	3.752	5117	- 23	-17.6	-42	1
Acetic acid	4.756	6486	- 92	-22.1	-37	1
Propionic acid	4.874	6647	- 163	-22.8	-38	1
Chloroacetic acid	2.861	3901	-1158	-17.0	-40	1
Glycolic acid	3.831	5225	175	-16.9	-39	1
Lactic acid	3.860	5267	- 99	-18.0	-41	1
Succinic acid, pK_1	4.207	5740	762	-16.7	-32	2
Succinic acid, pK_2	5.636	7693	- 108	-26.1	-52	2
Benzoic acid	4.202	5732	104	-18.9	-48	3
Phosphoric acid and derivatives						
Phosphoric acid, pK_1	2.148	2930	-1828	-16.0	-37	4
Phosphoric acid, pK_2	7.198	9823	987	-29.6	-54	5
Glycerol 2-phosphoric acid, pK_1	1.335	1820	-2893	-15.8	-78	6
Glycerol 2-phosphoric acid, pK_2	6.650	9069	- 412	-31.8	-54	6
Glucose 1-phosphoric acid, pK_2	6.504	8870	- 431	-31.2	-47	7
Ammonium ion and derivatives						
Ammonium ion	9.245	12614	12480	- 0.4	0	8, 9
Methylammonium ion	10.615	14484	13088	- 4.7	8	9
Dimethylammonium ion	10.765	14687	11859	- 9.5	23	9
Trimethylammonium ion	9.791	13358	8815	-15.2	44	9
Ethanolammonium ion	9.498	12958	12080	- 0.6	- 1	10
Tris(hydroxymethyl)aminomethane	8.076	11018	10900	- 0.3	—	11
Amino acids						
Glycine, pK_1	2.350	3205	1156	- 6.9	-32	12
Glycine, pK_2	9.780	13340	10550	- 9.4	-12	12
α -Alanine, pK_1	2.348	3203	773	- 8.2	-37	13
α -Alanine, pK_2	9.867	13458	10980	- 8.3	-17	13
α -Amino- <i>n</i> -butyric acid, pK_1	2.284	3123	298	- 9.5	-34	13
α -Amino- <i>n</i> -butyric acid, pK_2	9.831	13408	10695	- 9.1	-17	13
α -Aminoisobutyric acid, pK_1	2.357	3215	492	- 9.1	-38	13
α -Aminoisobutyric acid, pK_2	10.206	13919	11531	- 8.0	-15	13
Serine, pK_1	2.186	2980	1366	- 5.4	-32	14
Serine, pK_2	9.205	12550	10405	- 7.2	- 2	14
Threonine, pK_1	2.096	2859	1180	- 5.6	-32	14
Threonine, pK_2	9.100	12410	9960	- 8.2	-15	14
Hydroxyproline, pK_1	1.815	2476	918	- 5.2	-34	14
Hydroxyproline, pK_2	9.660	13160	9835	-12.7	-18	14
Proline, pK_1	1.970	2663	342	- 7.8	-38	14
Proline, pK_2	10.640	14510	10310	-14.1	-12	14

TABLE III (Continued)

Substance	pK_A	ΔF°	ΔH°	ΔS°	ΔC_p°	Ref.
β -Alanine, pK_1	3.551	4845	1179	-12.3	-30	15
β -Alanine, pK_2	10.235	13963	12570	-4.7	-3	15
γ -Aminobutyric acid, pK_1	4.031	5500	405	-17.1	-34	16
γ -Aminobutyric acid, pK_2	10.556	14400	12070	-7.8	-5	16
ϵ -Aminocaproic acid, pK_1	4.373	5965	-8	-20.0	-40	17
ϵ -Aminocaproic acid, pK_2	10.804	14740	13560	-4.0	8	17
Glycylglycine, pK_1	3.148	4140	862	-12.9	-40	17
Glycylglycine, pK_2	8.252	11260	10600	-2.0	-10	17
Aspartic acid, pK_1	1.995	2720	1783	-3.1	-29	17
Aspartic acid, pK_2	3.910	5335	1110	-14.2	-33	17
Aspartic acid, pK_3	10.006	13650	9025	-15.5	-21	17
Carbonic acid, pK_1	6.352	8666	2240	-21.6	-90	18
Carbonic acid, pK_2	10.329	14092	3603	-35.2	-65	19

All values are given on the molality scale (concentrations as moles per kilogram of water), but at 25° values on the molarity scale (moles per liter of solution) differ by only about 0.001 in pK_A . See text. ΔF° and ΔH° in cal. mole⁻¹; ΔS° and ΔC_p° in cal. deg⁻¹ mole⁻¹.

REFERENCES FOR TABLE III

1. H. S. Harned and B. B. Owen, "Physical Chemistry of Electrolytic Solutions," 2nd ed., p. 514, Reinhold Publishing Corp., New York, 1950.
2. R. G. Bates and G. Pinching, *J. Research Natl. Bur. Standards* **45**, 322, 444 (1950).
3. A. V. Jones and H. N. Parton, *Trans. Faraday Soc.* **48**, 8 (1952).
4. R. G. Bates, *J. Research Natl. Bur. Standards* **47**, 127 (1951).
5. R. G. Bates and S. F. Acree, *J. Research Natl. Bur. Standards* **30**, 127 (1943).
6. J. H. Ashby, E. M. Crook, and S. P. Datta, *Biochem J.* **56**, 198 (1954).
7. J. H. Ashby, H. B. Clarke, E. M. Crook, and S. P. Datta, *Biochem. J.* **59**, 203 (1955).
8. R. G. Bates and G. D. Pinching, *J. Research Natl. Bur. Standards* **42**, 149 (1949).
9. D. H. Everett and W. F. K. Wynne-Jones, *Trans. Faraday Soc.* **35**, 1380 (1939); *Proc. Roy. Soc.* **177A**, 499 (1940-41).
10. R. G. Bates and G. Pinching, *J. Research Natl. Bur. Standards* **46**, 349 (1951).
11. R. G. Bates and G. Pinching, *J. Research Natl. Bur. Standards* **43**, 519 (1949).
12. B. B. Owen, *J. Am. Chem. Soc.* **56**, 24 (1934); E. J. King, *ibid.* **73**, 155 (1951).
13. P. K. Smith, A. C. Taylor, and E. R. B. Smith, *J. Biol. Chem.* **122**, 109 (1937).
This paper also contains data for α -amino-*n*-valeric acid, α -amino-*n*-caproic acid, valine, leucine, and isoleucine, all of which are closely similar to the values for glycine.
14. P. K. Smith, A. T. Gorham, and E. R. B. Smith, *J. Biol. Chem.* **144**, 737 (1942).
15. M. May and W. A. Felsing, *J. Am. Chem. Soc.* **73**, 406 (1951).
16. E. J. King, *J. Am. Chem. Soc.* **73**, 155 (1951).
17. P. K. Smith and E. R. B. Smith, *J. Biol. Chem.* **146**, 187 (1942).
18. H. S. Harned and R. Davis, *J. Am. Chem. Soc.* **65**, 2030 (1943); see also R. Näsänen, *Acta Chem. Scand.* **1**, 204 (1947). For the effect of sodium chloride concentration on the first ionization of carbonic acid, see H. S. Harned and F. T. Bonner, *J. Am. Chem. Soc.* **67**, 1026 (1945).
19. H. S. Harned and S. R. Scholes, *J. Am. Chem. Soc.* **63**, 1706 (1941).

so that K_1 may be written (omitting the water involved, as usual in writing such equations)

$$K_1 = \frac{a_{\text{H}} a_{\text{HCO}_3}}{a_{\text{CO}_2}} = \frac{(\text{H}^+)(\text{HCO}_3^-)}{[(\text{H}_2\text{CO}_3) + (\text{CO}_2)]} \frac{\gamma_{\text{H}} \gamma_{\text{HCO}_3}}{\gamma_{\text{CO}_2}} \quad (88)$$

Here a_{CO_2} is so defined that it becomes equal to the total concentration of $\text{CO}_2 + \text{H}_2\text{CO}_3$ as the concentration approaches zero. The evaluation of the ratio of (H_2CO_3) to (CO_2) in solution is considered later, in Chapter 10. For carbonic acid pK_2 is defined by

$$K_2 = \frac{a_{\text{H}} a_{\text{CO}_3}}{a_{\text{HCO}_3}} = \frac{(\text{H}^+)(\text{CO}_3^{--})}{(\text{HCO}_3^-)} \frac{\gamma_{\text{H}} \gamma_{\text{CO}_3}}{\gamma_{\text{HCO}_3}} \quad (89)$$

The physical chemistry of carbonic acid is discussed in Chapter 10 and also in the last part of Volume II in connection with the physical chemistry of blood.

Some important regularities emerge from the data of Table III. The pK_A values for any homologous series, such as the fatty acids, are in general independent of the length of an alkyl group attached to the acidic group. There is, however, a marked change of the pK_A value involving an increase of approximately 1 pH unit on going from formic to acetic acid. The higher members of the fatty acid series, however, are almost identical with acetic acid; in fact the pK_A value for almost all of them lies between the values given for acetic and for propionic acids. The introduction of a methyl group into ammonium ion increases pK_A by 1.4 units at 25°. Here the introduced methyl group is considerably closer to the ionizing hydrogen than it is in acetic acid, and it is not surprising to find that its effect is greater. The higher straight-chain alkylammonium ions have pK_A values generally within 0.1 of those recorded in Table III for the methyl- or dimethylammonium ions.

In this preliminary interpretation of the data of Table III we have described the ionization of dibasic acids as occurring in two sharply distinct steps. This is very nearly true when the two pK values involved are separated by 4 or more pH units. When they are closer than this, however, as for instance in succinic acid, then there is some overlapping of the two stages of ionization and a more precise analysis is required. Indeed for a fundamental understanding of the data on polybasic acids in general such a discussion is essential and is provided in Chapter 9.

In discussing the ionization of the monoaminomonocarboxylic amino acids we have assumed that the isoelectric form that carries zero net charge is the dipolar ion and not the isomeric uncharged molecule. The discussion of the dielectric properties and the electrostatic interactions of the amino acids and peptides, already given in Chapters 5 and 6, pro-

vides abundant justification for this assumption. The evidence from ionization constants themselves, however, also provides quite conclusive evidence for the dipolar ion structure; indeed historically it was on this evidence that the argument for the dipolar ion structure of amino acids and peptides was first based. The details of the argument are given later in the discussion of polybasic acids; we shall simply assume the assignment of pK_A values here. The two pK_A values of the simple amino acids are so widely separated that for most purposes, without sensible error, we can consider pK_1 as referring to the ionization of the carboxyl group and pK_2 as referring to the ionization of the ammonium group. Comparison of the ionization of the carboxyl group in the glycine cation ($+H_3N \cdot CH_2 \cdot COOH$) and of propionic acid ($H_3C \cdot CH_2 \cdot COOH$), which are identical in electronic configuration, reveals the influence of the adjoining positive charge in increasing the acid strength of the carboxyl group—from 4.87 to 2.35 in pK , or 3240 calories in ΔF° . As the distance between the carboxyl and the amino group is increased—compare the values for β -alanine, γ -aminobutyric, and ϵ -aminocaproic acids—the influence of the positive charge progressively decreases, as might be expected. We return later to the calculation of the relation between the pK value and the distance between the charged group and the ionizing group.

Polar but uncharged substances adjoining the ionizing group have also a very marked effect on the pK values. Chloroacetic acid is nearly one hundred times as strong an acid as acetic. Glycolic and lactic acids, with a hydroxyl group in the α -position, are nearly ten times as strong.

The values of ΔH° show close relation to structure. For all the carboxylic acids these values are relatively small—generally less than 2 kcal/mole, many of them much less than 1. These values change rapidly with temperature, because of the large negative value of ΔC_p° for acids of this charge type. The ΔH° values for the phosphoric acids are somewhat smaller, although the values for the first ionization of phosphoric acid and for glycerol 2-phosphoric acid are numerically larger and more negative than for any of the carboxylic acids. On the other hand ΔH° values for the ammonium ions are very large, of the order of 10 to 12 kcal/mole. The corresponding differences appear also in the ionization of the amino acids and peptides, the ΔH° values associated with pK_1 being almost always numerically below 2 kcal/mole, as would be expected for the ionization of a carboxyl group, whereas those associated with pK_2 are 9 to 12 kcal/mole or more, which is characteristic for the ionization of the ammonium group of the amino acids.

In some respects it is more illuminating to consider the entropy of ionization (ΔS°) than the heat of ionization, since the latter is determined if the standard free energy and entropy of ionization are known, and

certain structural regularities are clearly associated with entropy effects. It is a general rule that acid-base equilibria of the class indicated in equations (82) and (83), in which an uncharged acid reacts with a neutral molecule, such as water, to produce an anion and a cation, are associated with a large negative value of ΔS° . This is not surprising. The reaction produces two ions where there were none before, and the intense electrostatic fields around these ions attract neighboring water dipoles, orient them, and compress them.¹⁰ This orientation and close packing inhibits the freedom of motion of the water dipoles, restricting the number of possible configurations which they can assume in solution, and thereby contributing a large negative term to the standard entropy of the reaction. Reactions of the charge type $\text{HA}^- + \text{H}_2\text{O} \rightleftharpoons \text{A}^{--} + \text{H}_3\text{O}^+$ generally give even larger negative values of ΔS° , as with the pK_2 values of succinic acid and the phosphoric acids in Table III. In the doubly charged phosphate ions the two negative charges are very close together, so that the large negative ΔS° values might be expected. In the doubly charged succinate ion, however, the two charges are much more widely separated, so it is perhaps somewhat unexpected that $-\Delta S^\circ$ for pK_2 should be considerably larger than for pK_1 .

The same electrostatic interactions, with the resultant "freezing" of the water dipoles in the neighborhood of the ions, also lead to a large decrease in the heat capacity of the system, the values of ΔC_p° being generally of the order of -40 to -50 cal deg⁻¹ mole⁻¹ in reactions such as (82) and (83).

On the other hand, reactions of the type indicated in equation (85), which simply involve the transfer of a proton from a substituted ammonium ion to a water molecule, involve no net formation of new ions.¹¹ The values of ΔS° and ΔC_p° for acid-base equilibria in reactions of this class are therefore numerically far less than those found for the reactions in which there is a net increase in the number of ions present as a result of proton transfer. The dimethyl- and trimethylammonium ions indeed are unusual, giving large *positive* values of ΔC_p° and moderately large negative values of ΔS° . Evidently, when these ions take part in proton transfer, there are peculiar effects, associated presumably with the fact that the surrounding methyl groups act as a shield interposed between

¹⁰ Reactions of this charge type are always associated with a decrease in the volume of the system; ΔV , per mole of (anion + cation) formed is near -10 cc for the ionization of a carboxylic acid, and -21 cc for the ionization of water.

¹¹ These are sometimes called "isoelectric reactions," which is an appropriate term, but the use of the word "isoelectric" in this connection should not be confused with its much more common and important use to denote the condition of zero electric mobility for amphoteric substances. This is discussed in detail in Chapter 9.

the charge on the central nitrogen atom and the water dipoles which are being oriented in the surrounding medium. A further study of these effects would be of great interest; but apart from this, if we consider only substances containing primary amino groups, the rule holds that the standard heat and entropy changes are very small for acid-base equilibria involving only charge transfer and not the formation of new charges.

The standard entropy and heat capacity changes associated with the pK_A values for the amino acids are interpretable along similar lines. Reactions of the type (86) for pK_1 involve the formation of two new charges, but only one of them is a free ion; the other is a negatively charged $-\text{COO}^-$ group which is part of a dipolar ion. The neighboring $-\text{NH}_3^+$ group, especially if it is in the α -position, partially neutralizes the electrostatic interactions between the COO^- group and the surrounding water dipoles, and the value of ΔS° , though negative, is numerically far smaller than for the fatty acids. If the charged groups are more widely separated, however, as with γ -aminobutyric and ϵ -aminocaproic acids, the effect of the $-\text{NH}_3^+$ group on ΔS° falls off rapidly, and the values obtained are close to those for the fatty acids. It should be noted that the ion-dipole interactions associated with the ΔS° values are short-range interactions; the charged $-\text{NH}_3^+$ group has relatively little effect on them unless it is in the α -position. On the other hand, the effect of the $-\text{NH}_3^+$ group on ΔF° for the carboxyl group ionization is due to long-range interactions between groups carrying net charges. Thus the pK_1 value, even for ϵ -aminocaproic acid, is significantly different from that for an unsubstituted fatty acid (see below).

The ΔC_p° values, for pK_1 of the α -amino acids, are negative and numerically almost as large as for the fatty acids. The adjoining $-\text{NH}_3^+$ group seems to have much less effect on ΔC_p° than on ΔS° . No ready explanation of this fact can yet be offered. Reaction (87), for pK_2 of the amino acids, is of the "isoelectric" type. No new charges are created, but a positive charge is removed from an $-\text{NH}_3^+$ group adjoining a negatively charged $-\text{COO}^-$ group and transferred to a water molecule. Thus the two charges, by becoming more widely separated, can exert a more powerful attraction on the surrounding water dipoles, and the moderately negative values of ΔS° and ΔC_p° for pK_2 of the α -amino acids are thus reasonable. As the separation between the charges in the dipolar ion increases, their mutual interaction rapidly falls off; the values of pK_2 of ϵ -aminocaproic acid are close to those typical of primary ammonium ions.

Effect of Neighboring Charged Groups and Dipoles on pK_A Values

The presence of a neighboring substituent group carrying a net charge, or of a dipolar group, may greatly alter the characteristic pK value of

any acidic group. Some striking examples are seen in Table III, particularly in the data for the amino acids. Some further examples are given in Tables IV and V, where the variation of pK with the number of carbon atoms separating the substituent group from the acidic group under consideration is shown. It will be noticed that, for dipolar substituents, there is a general parallelism between the dipole moment of the substituent

TABLE IV^a

INFLUENCE OF SUBSTITUENTS (R) IN DIFFERENT POSITIONS ON THE DISSOCIATION (pK') OF THE CARBOXYL GROUP IN $R \cdot (CH_2)_nCOOH$ AT 25°

Substituent (R)	Dipole moment of CH_3R (Debye units)	Position			
		α ($n = 1$)	β ($n = 2$)	γ ($n = 3$)	δ ($n = 4$)
CH_3	0	4.87	4.83	4.80	4.85
$CH_2=CH$	0.34	4.42			
C_6H_5	0.39	4.26			
HO	1.65	3.82			
HS	1.39	3.60			
COOR'	1.7	3.34	4.52		4.60
I	1.6	3.15	4.05	4.64	4.77
Br	1.8	2.86	4.01	4.58	4.72
Cl	1.8-1.9	2.81	4.07	4.52	4.69
COOH	1.7	2.92	4.24	4.36	4.42
O_2N	3.0-3.8		3.79		
$N \equiv C$	3.1-3.5	2.44			
NH_3^+		2.35	3.55	4.03	4.21

^a From Cohn and Edsall (1943), Chapter 5.

Values for the acidity effect of the $-COOH$ group from R. Gane and C. K. Ingold, *J. Chem. Soc.* 2158 (1931); G. Schwarzenbach, *Helv. Chim. Acta* **16**, 522 (1933); for the $-NH_3^+$ group, Table III above. Other values of pK from Landolt-Börnstein and International Critical Tables. Values of dipole moments are taken from the table in *Trans. Faraday Soc.* **30**, 904ff. (1934). Two substituents which do not fit into the table should be noted: (1) the $C=O$ group: pyruvic acid ($CH_3CO \cdot COOH$) has a pK' of 2.49; the dipole moment of acetone is 2.7D; (2) the $C \equiv C$ triple bond: tetrolic acid ($CH_3-C \equiv C \cdot COOH$) has $pK' = 2.60$.

group and its effect on ΔpK_A , which is the displacement of pK_A for the substituted acid as compared with the unsubstituted fatty acid or ammonium ion. The correlation is not by any means perfect; thus the $-SH$ group has a lower dipole moment than the $-OH$ group, but a somewhat greater effect on ΔpK_A . The effect of a charged $-NH_3^+$ group on the pK value of a $-COOH$ group or another $-NH_3^+$ group extends over a much larger distance than the effect of a dipolar substituent. For instance, the pK_A of ϵ -aminocaproic acid is 4.37 ($\Delta pK_A = 0.50$), whereas no dipolar substituent, even in the γ -position—two carbon atoms closer—

gives as large a ΔpK_A value as this. This is of course to be expected on electrostatic grounds (Chapters 5 and 6), since the electric field intensity due to a dipole falls off much more rapidly with distance than that due to an ion.

TABLE V
INFLUENCE OF SUBSTITUENTS (R) ON THE ACIDITY (pK') OF THE AMMONIUM GROUP IN $R(CH_2)_nNH_3^+$ AT 20° OR 25°

Substituent (R)	Dipole moment of CH_3R (Debye units)	Position of substituent			
		α ($n = 1$)	β ($n = 2$)	γ ($n = 3$)	δ ($n = 4$)
CH_3-	0	10.66	10.59	10.68	10.70
$CH_2=CH-$	0.34	9.76			
C_6H_5-	0.39	9.38			
$HO-$	1.65		9.48		
H_2N-	1.23		9.98	10.62	10.86
$-COOR'$	1.7	7.75	9.13	9.71	10.15
$-COO^-$		9.72	10.19	10.40	10.69
$-NH_3^+$			6.98	8.58	9.32

Values for the charged $-NH_3^+$ group (at 20°) from G. Schwarzenbach, *Helv. Chim. Acta* **16**, 522 (1933); for the $-COOR'$ group (at 25°) from J. T. Edsall and M. H. Blanchard, *J. Am. Chem. Soc.* **55**, 2337 (1933), and from A. Neuberger, *Proc. Roy. Soc. A* **158**, 68 (1937). Other values of pK from Landolt-Börnstein and International Critical Tables. Dipole moments, see Chapter 6.

For the $^+H_3N \cdot NH_3^+$ ion, G. Schwarzenbach [*Helv. Chim. Acta* **19**, 178 (1936)] gives $pK_1 = -0.88$; for $^+H_3N \cdot NH_2$, $pK(pK_2) = 8.10$ at 20°.

The pK values for the diammonium ions (substituents $-NH_3^+$ and $-NH_2$) are uncorrected for the statistical effect (see discussion in Chapter 9). The pK of an individual NH_3^+ group in $^+H_3N \cdot (CH_2)_n \cdot NH_3^+$ is obtained by adding 0.30 to the pK values given above for the substituent $-NH_3^+$. Likewise the pK of an individual $-NH_3^+$ group in $H_2N \cdot (CH_2)_n \cdot NH_3^+$ is given by subtracting 0.30 from the values given above for the substituent H_2N- .

From Cohn and Edsall (1943), Chapter 5.

The first quantitative formulation of the electrostatic effect of a charged group on the acidity of a neighboring group was given by Bjerrum (1923). If the group of charge $Z\epsilon$ —where Z is the valence of the group, and ϵ is the proton unit of charge—is regarded as immersed in a medium of dielectric constant D , and if it is at a distance r from the acidic group, then the contribution of the charged group to the electrical potential at the acidic group is

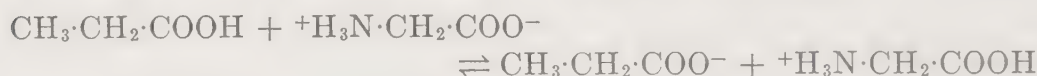
$$\psi = \frac{Z\epsilon}{Dr} \quad (90)$$

Bjerrum set D equal to the dielectric constant of the pure solvent. The

electrical work done under the influence of the potential, ψ , in removing a proton from the acidic group at r to infinity is, per mole,

$$\Delta W = N\epsilon\psi = \frac{NZ\epsilon^2}{Dr} \quad (91)$$

If now we compare the free energy of ionization of an acid containing a charged substituent group with that of a similar acid containing no such substituent, the difference in ΔF° for the two acids may be set equal to ΔW . Compare, for instance, the glycine cation ($^+\text{H}_3\text{N}\cdot\text{CH}_2\cdot\text{COOH}$), $pK_1 = 2.350$ at 25° , with propionic acid ($\text{CH}_3\cdot\text{CH}_2\cdot\text{COOH}$), $pK_A = 4.874$ at 25° . Both acids have the same electronic configuration and approximately the same size. The work done in the proton transfer:



should be equal to ΔW , and the equilibrium constant of this reaction is equal (see equation 41) to K_A (propionic acid)/ K_A (glycine). Thus we may write, making use of (91):

$$\begin{aligned} \Delta W &= \Delta F^\circ (\text{propionic acid}) - \Delta F^\circ (\text{glycine}) \\ &= 2.303RT[(pK (\text{propionic acid}) - pK (\text{glycine}))] = \frac{N\epsilon^2}{RTDr} \quad (92) \end{aligned}$$

Taking the numerical values $\epsilon = 4.80 \times 10^{-10}$ esu, $N = 6.02 \times 10^{23}$, $R = 8.31 \times 10^7$ erg deg $^{-1}$ mole $^{-1}$, $T = 298.1^\circ$, and $D = 78.5$, and expressing r in angstroms, this becomes

$$r = \frac{3.08}{\Delta pK} = \frac{3.09}{4.874 - 2.350} = 1.23 \text{ \AA} \quad (93)$$

This value is obviously much too low, being less than the distance between two bonded carbon atoms. When the substituent group, however, is further removed from the acidic group, the values of r calculated from Bjerrum's formula accord much better with those calculated from known interatomic distances in the molecule (Neuberger, 1937). Similar estimates of r may be obtained by comparing the dissociation of the ammonium group in an amino acid (pK_2) with the dissociation of the same group in a corresponding amino acid ester.

It is obvious that Bjerrum's treatment can be only an approximation. The solvent is not a continuous medium; the solute molecules must form cavities of considerable size, and the dielectric constant of such a cavity is far lower than that of a medium like water. A large part of the electrical effect of the substituent must be transmitted through the molecule, that is, through a medium of low dielectric constant. Hence the effect of the

substituent should be *greater* than that calculated for a medium of the dielectric constant of the solvent; thus the distances calculated from equation (91) naturally turn out to be too low. This is particularly true when substituent and ionizing groups are very near; as the separation increases, more of the electrical effect is transmitted through the solvent.

These considerations formed the basis for a revision of Bjerrum's hypothesis by Kirkwood and Westheimer (1938), which yields far more satisfactory results than the theory in its original form. In this treatment the solute molecules are considered as cavities of dielectric constant D_i in the solvent. The value of D_i is taken as 2, near that of the liquid paraffin hydrocarbons; but a moderate change in the value of D_i makes very little difference in the outcome of the calculations. Kirkwood and Westheimer developed this treatment for two models: (1) a spherical molecule of radius b , with an arbitrary charge distribution, immersed in a solvent of dielectric constant D ; (2) an ellipsoid of revolution, of any eccentricity—the charged substituent and the ionizing group are generally taken as located at the foci of the ellipsoid.

It is found that, for either of these models, the Bjerrum equation (91) still holds, if the dielectric constant of the solvent is replaced by an "effective dielectric constant," D_E . For the spherical model, D_E is a function of the position of the substituent and of the ionizing group within the sphere; for the ellipsoidal model, it is a function of the eccentricity of the ellipse. The specific values of D_E applicable to any particular structure have been tabulated by Kirkwood and Westheimer. In all cases, their treatment leads to reasonable values of r entirely compatible with known data on interatomic distances and molecular configurations. Westheimer and Shookhoff (1939) have calculated the interprotonic distances in a large number of symmetrical dicarboxylic acids from the equation, similar to (91):

$$\Delta pK = \log \frac{K_1}{K_2} - \log 4 = \frac{N\epsilon^2}{2.303RTD_E r} \quad (94)$$

Here the term $(-\log 4)$ must be inserted, because for acids of this type $K_1 = 4K_2$ in the limiting case when the acidic groups are so far apart that neither influences the dissociation constant of the other. (The basis for this "statistical factor" of 4 is discussed in detail in Chapter 9.) Their treatment gives for oxalic acid ($\Delta pK = 2.36$), for instance, 3.83 Å as the distance between the two acidic protons—a value entirely in harmony with the dimensions of the molecule as known from crystal structure determination—whereas the simple Bjerrum treatment gives the impossibly low value of 0.91 Å. For malonic acid ($\Delta pK = 2.26$) Westheimer and Shookhoff find $r = 4.10$ Å, where the earlier treatment gives

1.36 Å. Diethylmalonic acid gives a much larger ΔpK value (4.48), but the calculated interprotonic distance (3.75 Å) is only slightly less than in malonic acid, since the presence of the two bulky ethyl groups increases the region of low dielectric constant adjoining the carboxyls and thereby enhances the electrostatic interactions, thus increasing ΔpK .

A similar treatment should be applicable to the diammonium ions (hydrazinium, ethylenediammonium, etc.) whose acidic constants have been determined by Schwarzenbach (1933, 1936) in water and alcohol-water mixtures. The aliphatic dimercaptans (Schwarzenbach and Epprecht, 1936), which have been studied in alcohol-water mixtures, should also furnish material for such calculations.

The same kind of electrostatic picture can be employed for a dipolar substituent. The effect of a dipole of moment μ in a homogeneous medium of dielectric constant D should be given by the equation

$$\Delta pK = \frac{N\epsilon\mu \cos \theta}{2.303RTDr^2} \quad (95)$$

where θ is the angle between the dipole and the line joining its center to the ionizable proton, and r is the distance between the two along this line. (See Chapter 5, equations 23 and 86.) Like Bjerrum's equation for charged substituents, (95) gives values of r which are much too low, if D is taken as the dielectric constant of the solvent, and if the dipole and the acidic group are close together. Kirkwood and Westheimer have treated the effect of dipolar substituents in a very similar manner to their treatment of charged substituents, but we shall not discuss their calculations here.

A charged or polar substituent sets up an inductive displacement of electronic charges within the molecule. In a $+C - Cl^-$ dipole, for instance, the carbon atom at the positive end of the dipole tends to attract electrons from the next atom adjoining it in the chain, and that atom in turn attracts electrons from its next neighbor, and so forth. This internally transmitted inductive effect must influence the pK value of an acidic group in another part of the molecule; and it may be difficult to disentangle this effect from the direct electrostatic field effect of a charged or polar substituent acting on the acidic group. In a saturated aliphatic chain, however, the inductive effect falls off very rapidly with increasing length of the chain, so that ΔpK_A for ϵ -aminocaproic acid, for instance, must be determined primarily by the electrostatic field set up by the charged $-NH_3^+$ group in the ϵ -position. An important discussion of the relative importance of the direct and the inductive effects has been given by Ingold (1953).

It is well known that aromatic ammonium groups, as in the anilinium

ion ($\text{C}_6\text{H}_5\text{NH}_3^+$, pK_A' 4.6), are far stronger acids, by a factor of the order of a million, than the corresponding aliphatic ammonium ions. Similarly the conjugate acid of pyridine (pK_A' 5.2) is about a million times as strong as that of piperidine (pK_A' 11.1). These differences have been interpreted in terms of resonance effects and are discussed elsewhere.¹²

Other Acidic Groups of Biochemical Interest

Many compounds containing acidic groups of great biochemical importance have not been studied in cells without liquid junction but have been studied in cells with liquid junction with either hydrogen or glass electrodes. The data obtained on such cells are of sufficient accuracy to furnish much important information, and a selection of values is given in Table VI. Some of these values have been corrected for ionic strength effects, so as to obtain a true limiting value for pK ; most are pK' values, which may differ by as much as 0.15 from the true pK values. For the major effects with which we are concerned, however, these differences are relatively trivial. The data deserve some comment.

PHENOLIC HYDROXYL GROUPS

The characteristic pK value of these groups at 25° is near 10. The substitution of two iodine atoms on the carbons immediately adjoining that to which the hydroxyl group is attached, as in 2,6-diiodotyrosine, lowers the pK value by nearly 3.5 units. This great decrease in ΔF° is accompanied by an almost exactly equal decrease in ΔH° , from 6000 to below 1000 cal mole⁻¹. The characteristic standard entropy of ionization is very nearly -26, in all the three cases for which this value is known (Table VI). As with the carboxylic acids, this acid-base equilibrium is of the type in which an anion and a cation are produced by the interaction of water and an uncharged group—here the phenolic group. In all such instances the value of ΔS° is found to be large and negative, and even larger for the phenols than for the carboxylic acids. It appears that the value of ΔS° is of the order of magnitude of -20 to -25 cal deg⁻¹ mole⁻¹ for any uncharged acid which reacts with water to give an anion and a cation, unless the process is modified by the effect of a charged group very close to the acidic group, or unless the acidic hydrogen is involved in some special type of hydrogen bonding.

The value of pK_2 for salicylic acid—reported as 13.4 at 18°—is much larger than that for any of the other phenolic groups recorded in Table VI. Here there is ample evidence from other sources that this unusual value

¹² See, for instance, Cohn and Edsall (1943), pp. 124-130; and Ingold (1953).

TABLE VI
 APPARENT DISSOCIATION CONSTANTS, EXPRESSED AS pK_A' VALUES FOR
 SUBSTANCES CONTAINING VARIOUS ACIDIC GROUPS OF BIOCHEMICAL
 INTEREST
 (Values for 25° unless otherwise indicated)

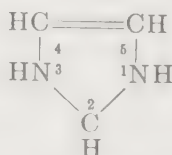
1. Substances Containing Aromatic (Phenolic) Hydroxyl Groups			
Substance	pK_A'	ΔH°	ΔS°
Phenol	9.98	6100	-25
Salicylic acid, pK_2 (—OH)	13.4		
<i>m</i> -Hydroxybenzoic acid, pK_2 (—OH)	9.99		
<i>p</i> -Hydroxybenzoic acid, pK_2 (—OH)	9.4		
Tyrosine, pK_1 (—COOH)	2.20		
Tyrosine, pK_2 (—NH ₃ ⁺)*	9.11		
Tyrosine, pK_3 (—OH)*	10.95	6000	-26
2,6-Diiodotyrosine, pK_1 (—COOH)	2.12	980	-7
2,6-Diiodotyrosine, pK_2 (—OH)	6.48	810	-27
2,6-Diiodotyrosine, pK_3 (—NH ₃ ⁺)	7.82	8790	-6

Data for the hydroxybenzoic acids from E. Larsson, *Z. inorg. allgem. Chem.* **183**, 30 (1929). Most of the other data are from E. J. Cohn and J. T. Edsall, "Proteins, Amino Acids and Peptides," Chapter 4, Reinhold Publishing Corp., New York, 1943. Larsson's values are for 18°.

Concerning the spectrophotometric determination of the pK value of the hydroxyl group of tyrosine, see J. L. Crammer and A. Neuberger, *Biochem J.* **37**, 302 (1943); D. Shugar, *ibid.* **52**, 142 (1952); C. Fromageot and G. Schnek, *Biochim. et Biophys. Acta* **6**, 113 (1950-51); C. Tanford and G. L. Roberts, Jr., *J. Am. Chem. Soc.* **74**, 2509 (1952). The ΔH° and ΔS° values for pK_3 of tyrosine are from the latter paper.

* The values of pK_2 and pK_3 for tyrosine are hybrid constants, each involving both the —NH₃⁺ and the —OH group. See the discussion in Chapter 9, p. 504 (footnote).

2. Substances Containing Imidazole (Im) Groups			
Substance	pK_A'	ΔH°	ΔS°
Imidazole	6.95	7700	-6
4-Methylimidazole	7.52	8600	-5
2,4-Dimethylimidazole	8.36	9200	-7
Histidine, pK_1 (—COOH)	1.82	1200	-4
Histidine, pK_2 (—Im)	6.00	6900	-4
Histidine, pK_3 (—NH ₃ ⁺)	9.17	9400	-10



The numbering of the atoms in the imidazole ring is shown above. Data from E. J. Cohn and J. T. Edsall, "Proteins, Amino Acids, and Peptides" (1943), Chapter 4, and for the imidazoles especially from A. H. M. Kirby and A. Neuberger, *Biochem. J.* **32**, 1146 (1938). ΔH° values for the first three compounds listed are from Y. Nozaki, F. R. N. Gurd, R. F. Chen, and J. T. Edsall, *J. Am. Chem. Soc.* **79**, 2123 (1957).

TABLE VI (Continued)

3. Substances Containing Sulfhydryl Groups

Substance	pK_A	Substance	pK_A
Thioglycolic acid, pK_1 (—COOH)	3.67	Cysteinylcysteine, pK_1	2.65
Thioglycolic acid, pK_2 (—SH)	10.31	(—COOH)	
Methylthioglycolate (—SH)	7.8	Cysteinylcysteine, pK_2^*	7.27
HO—CH ₂ ·CH ₂ —SH	9.5	Cysteinylcysteine, pK_3^*	9.35
⁺ H ₃ N·CH ₂ ·CH ₂ ·SH (pK_1)*	8.6	Cysteinylcysteine, pK_4^*	10.85
⁺ H ₃ N·CH ₂ ·CH ₂ ·SH (pK_2)*	10.75	Glutathione, pK_1 (COOH)	2.12
Cysteine, pK_1 (—COOH)	1.8	Glutathione, pK_2 (COOH)	3.59
Cysteine, pK_2^*	8.3	Glutathione, pK_3^*	8.75
Cysteine, pK_3^*	10.8	Glutathione, pK_4^*	9.65
Cysteine betaine, pK_2 (SH), (N-trimethylcysteine)	8.6		

ΔH° is given as 6900 cal/mole for the ionization of the —SH group in thioglycolic acid; hence $\Delta S^\circ = -24$ cal deg⁻¹ mole⁻¹ for this ionization (see reference to Benesch and Benesch, below).

Value for thioglycolic acid (pK_1) from E. Larsson, *Z. anorg. allgem. Chem.* **172**, 375 (1928); pK_2 from R. E. Benesch and R. Benesch, *J. Am. Chem. Soc.* **77**, 5877 (1955); see also R. K. Cannan and B. C. J. G. Knight *Biochem. J.* **21**, 1384 (1927). Other values from J. Neilands as reported by M. Calvin in "Glutathione: A Symposium" (S. Colowick *et al.*, eds.), p. 9, Academic Press, New York, 1954; and from E. J. Cohn and J. T. Edsall, "Proteins, Amino Acids and Peptides," p. 85, Reinhold Publishing Corp., New York, 1943. Values for pK_2 , pK_3 , and pK_4 of glutathione are taken from N. C. Li, O. Gawron, and G. Bascuas, *J. Am. Chem. Soc.* **76**, 225 (1954). Values of pK marked* are hybrid constants, involving both the —NH₃ and the —SH groups. [See L. R. Rykkan and C. L. A. Schmidt, *Arch. Biochem.* **5**, 89 (1944).] The interpretation of such hybrid constants is discussed later. See Chapter 9, Tables II, III and VI.

4. Lysine, Arginine, Guanidine, and Creatine

Substance	pK_A
Guanidine	ca. 14
Arginine, pK_1 (—COOH)	1.81
Arginine, pK_2 (—NH ₃ ⁺)	9.01
Arginine, pK_3 (Guan)	ca. 12.5
Creatine, pK_1 (—COOH)	2.62
Lysine, pK_1 (—COOH)	2.16
Lysine, pK_2 (α-NH ₃ ⁺)	9.18
Lysine, pK_3 (ε-NH ₃ ⁺)	10.79
Lysyl-lysine (LL) pK_1 (—COOH)	3.01
Lysyl-lysine, pK_2 (α-NH ₃ ⁺)	7.53
Lysyl-lysine, pK_3 (ε-NH ₃ ⁺)	10.05
Lysyl-lysine, pK_4 (ε-NH ₃ ⁺)	11.01

pK_1 for creatine from R. K. Cannan and H. Shore, *Biochem. J.* **22**, 920 (1928); for numerous guanidine derivatives, not listed here, see T. L. Davis and R. C. Elderfield, *J. Am. Chem. Soc.* **54**, 1499 (1932). For other references see Section 5 below.

TABLE VI (Continued)

5. Some Glycyl and Alanyl Peptides and Substances Containing Peptide Linkages		
Substance	pK_1 ($-\text{COOH}$)	pK_2 ($\alpha\text{-NH}_3^+$)
Acetylglycine	3.60	
Gly-Gly-Gly	3.26	7.91
Gly-Gly-Gly-Gly	3.05	7.75
Gly-Ala	3.17	8.23
Ala-Gly	3.17	8.18
Gly-Ala-Ala (LL)	3.38	8.10
Gly-Ala-Ala (LD)	3.30	8.17
Ala-Ala-Ala-Ala (LLLL)	3.42	7.94
Ala-Ala-Ala-Ala (LLDL)	3.24	7.93
Ala-Ala-Ala-Ala (LDLL)	3.22	7.99
Ala-Ala-Ala-Ala (DLLL)	3.42	7.99

Data from E. J. Cohn and J. T. Edsall, "Proteins, Amino Acids and Peptides," p. 85, Reinhold Publishing Corp., New York, 1943; and from E. Ellenbogen, *J. Am. Chem. Soc.* **74**, 5198 (1952). Abbreviations for peptides as in Chapter III. Notation for L or D configuration of successive amino acid residues in a sequence follows symbols for residues (see Ellenbogen's paper for further information).

Data for many other peptides are listed in the references given above.

6. Some Effects of the Keto Group and the Peptide Linkage on the Acidity of Neighboring Groups

A. Effect of Peptide Linkage on $-\text{NH}_3^+$ groups

	pK (NH_3^+) (amino acid or peptide)	pK' (NH_3^+) (ethyl ester)
$^+\text{H}_3\text{N}\cdot\text{CH}_2\cdot\text{CO}\cdot\text{NH}\cdot\text{CH}_2\text{COO}^-$	8.25	7.75
$^+\text{H}_3\text{N}\cdot\text{CH}_2\cdot\text{CH}_2\cdot\text{CH}_2\cdot\text{CH}_2\cdot\text{COO}^-$	10.76	10.15
ΔpK	2.51	2.40

B. Effect of Peptide Linkage on $-\text{COOH}$ Groups

	pK (COOH)	pK (COOH)
$\text{R}\cdot\text{CH}_2\cdot\text{CO}\cdot\text{NH}\cdot\text{CH}_2\cdot\text{COOH}$	3.60 (R = H)	3.15 (R = NH_3^+)
$\text{R}\cdot\text{CH}_2\cdot\text{CH}_2\cdot\text{CH}_2\cdot\text{CH}_2\cdot\text{COOH}$	4.80	4.27
ΔpK	1.20	1.12

C. Effect of Keto Group on Ionization of the $-\text{COOH}$ Group

		pK'
Pyruvic acid	$\text{CH}_3\text{CO}\cdot\text{COOH}$	2.5
Acetoacetic acid	$\text{CH}_3\text{CO}\cdot\text{CH}_2\cdot\text{COOH}$	3.58
Levulinic acid	$\text{CH}_3\text{CO}\cdot\text{CH}_2\cdot\text{CH}_2\cdot\text{COOH}$	4.59
Valeric acid	$\text{CH}_3\text{CH}_2\cdot\text{CH}_2\cdot\text{CH}_2\cdot\text{COOH}$	4.80

For references see other parts of this table, and Table III; also E. J. Cohn and J. T. Edsall, "Proteins, Amino Acids and Peptides," Chapters 4 and 5, Reinhold Publishing Corp., New York, 1943.

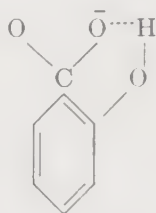
TABLE VI (Continued)

7. Some Purines, Pyrimidines, Nucleosides, and Nucleotides

	pK_A' ($-\text{NH}_3$)	pK_A' (sec. phosphate)	pK_A' ($-\text{OH}$ or $-\text{NH}$)
Uracil	—	—	9.45
Thymine	—	—	9.94
Cytosine	4.60	—	12.16
Adenine	4.15	—	9.80
Guanine	3.3	—	9.20, 12.3
Adenosine	3.3	—	12.5
Guanosine	1.6	—	9.2, 12.3
Uridine	—	—	9.2, 12.3
Adenosine-5'-phosphate	3.74	6.05	
Adenosine diphosphate	3.95	6.26	
Adenosine triphosphate	4.00	6.48	

The last three pairs of values are from R. A. Alberty, R. M. Smith, and R. M. Bock, *J. Biol. Chem.* **193**, 425 (1951); others from P. A. Levene *et al.*, *ibid.* **65**, 519 (1925); **70**, 229, 243 (1926); and H. F. W. Taylor, *J. Chem. Soc.* 765 (1948); see also *Nature* **164**, 750 (1949). Extensive tabulations of data are given by D. O. Jordan in "The Nucleic Acids" (E. Chargaff and J. N. Davidson, eds.), Vol. I, Chapter 13, Academic Press, New York, 1955. See also Chapters 6 and 8 of the same book. Adenosine-5-phosphate (muscle adenylic acid) contains one primary phosphoric acid group, adenosine diphosphate contains two such groups, and adenosine triphosphate three. Their pK values lie in the strongly acid region; the value for adenylic acid was reported as 0.89 by Levene and Simms; the others have apparently not yet been reported.

is due to hydrogen bonding between the $-\text{OH}$ group and the adjoining $-\text{COO}^-$ group in the *ortho* position. The strong attraction exerted on the



acidic hydrogen by the negative charge on the closely adjoining carboxylate ion makes its removal by a base far more difficult. In contrast, *m*- and *p*-hydroxybenzoic acids show typical phenolic pK values. The unusual behavior of the hydroxyl group of salicylic acid finds a parallel in the titration curves of a number of proteins, such as ovalbumin (Crammer and Neuberger, 1943) and ribonuclease (Shugar, 1952), in which some or all of the phenolic groups of tyrosine residues do not ionize except at high pH values, commonly above 12. In such proteins, when ionization occurs, there may be irreversible alteration and the protein is denatured. Pre-

sumably the hydrogen bonding here involves the —OH groups of certain tyrosine residues, and perhaps also the ionized carboxyl groups of aspartic or glutamic acid residues in other portions of the protein molecule, which are geometrically so placed that bonding can occur. The breaking of such bonds would then represent one aspect of the disorganization of the highly organized structure of the native protein, which occurs during protein denaturation. In the case of serum albumin (Tanford and Roberts, 1952) there is evidence that the phenolic groups in the native protein are involved in hydrogen bonding, but the bonds are broken and re-formed with relative ease, and apparently reversibly, if the protein is not exposed to too-alkaline a pH . (See the further discussion in Chapter 9.)

IMIDAZOLE GROUPS

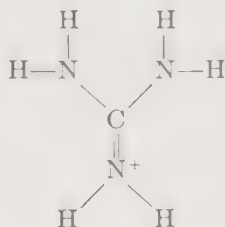
The imidazole groups of histidine residues in proteins play a part of major importance in physiological buffer systems, since their characteristic pK values are close to 7. Values for several simple compounds containing imidazole groups are listed in Table VI. The increasing basicity of the imidazoles with increasing numbers of methyl groups attached to the imidazole ring carbon atoms is apparent. Qualitatively, at least, this is in accord with the effects to be expected on replacing a hydrogen atom by a more polarizable alkyl group, which acts as a donor of electrons to the ring, and increases the basicity of the ring nitrogens. The pK_A value for the imidazole group in histidine, which has the side chain attached in the same position as the methyl group in 4-methylimidazole, is 1.5 pH units more acid than the latter, presumably owing primarily to the effect of the charged ammonium group. The standard entropies of ionization are negative but small, of the same order of magnitude as the value for the methylammonium ion (Table III).

SULFHYDRYL GROUPS

These groups are of major importance in oxidation-reduction reactions and are essential to the functioning of many enzymes and other proteins. Since the alkyl mercaptans are insoluble in water, there are no data directly available to determine the pK_A values of —SH groups attached only to an alkyl chain. All the —SH compounds listed in Table VI contain other polar groups which render them water-soluble. We may note, however, that the pK_A value of $HO\cdot CH_2CH_2SH$ is 9.5, which is almost identical with the value of 9.48 for the ammonium group in the exactly analogous compound $HO\cdot CH_2CH_2NH_3^+$. Since an unsubstituted ammonium group in an alkylammonium ion has a pK_A near 10.7 at 25°, this suggests that the value for an unsubstituted alkyl mercaptan should be very similar.

In its effect on the acidity of another neighboring acid group, the sulfhydryl group is somewhat more powerful than a hydroxyl group; compare glycolic acid (pK 3.83) with thioglycolic acid (pK 3.67). The interpretation of the pK values for the $-SH$ and the $-NH_3^+$ groups of cysteine involves the evaluation of four "microscopic" constants for the individual groups; this problem is considered in detail in Chapter 9.

THE GUANIDINIUM GROUP



The guanidinium ion is a symmetrical structure, with the three nitrogen atoms placed in the form of an equilateral triangle, all equidistant from the central carbon atom and coplanar with it. The structure involves resonance between the formula shown, and two other equivalent formulas in which the double bond is located on one of the other C—N bonds. Thus all the C—N bonds have one-third double-bond character, and this should be sufficient to fix the position of the hydrogen atoms so that they are coplanar with the carbon and nitrogen atoms, as in the closely analogous structure of urea. If a proton is removed from the guanidinium ion, the conjugate base, guanidine, no longer possesses the symmetrically resonating structure of the ion. This stabilizes the latter, so that the removal of a proton from the guanidinium ion requires a large amount of work, and the pK_A value of the guanidinium ion is about 14. The simple N-alkyl, N,N-dialkyl, and N,N',N''-trialkyl derivatives are similar in their pK_A values; but the disubstituted derivatives in which there is one alkyl group on each of two of the three nitrogens are considerably more acidic, with a pK_A value of the order of 12.3. The guanidinium group of arginine is intermediate, with pK near 12.5 at 25°. Thus in the titration of a protein, in which pH values as high as 13 are seldom attained, and even values above 12 are measured only with some technical difficulty, it is fairly certain that many of the guanidinium groups of the arginine side chains are still positively charged even at the most alkaline pH values attained. In contrast, it is relatively easy, on the acid side of the titration curve, to obtain quite accurate measurements at pH values of 1.5 to 2, at which virtually all the negative charges of the carboxyl, phenolic, and sulfhydryl groups have been neutralized by proton binding.

THE PEPTIDE LINKAGE

The —CO·NH— group has a powerful effect in increasing the acidity of a neighboring group. Many peptides have been studied, only a few of which are listed in Section 5 of Table VI. It is clear, however, that the terminal α -ammonium group at the end of a long peptide chain has a characteristic pK' value below 8; values for cystine peptides (not shown here) show that the presence of a neighboring disulfide linkage can depress these values still further, to 6.5 to 7. Thus terminal α -amino groups, at the ends of peptide chains in proteins, can contribute to the buffering power of protein solutions in the physiological pH range.

Section 6 of Table VI shows the effect of replacing a $\text{—CH}_2\text{CH}_2\text{—}$ group by a —CO·NH— group, in an aliphatic chain, on the ionization of a neighboring —NH_3^+ or —COO^- group. The effect on the former is far more profound, as might be expected, since the ionizing proton is considerably closer to the peptide linkage in the former case. The large effect of an α -keto group on the ionization of a carboxyl group, and the decreasing effect of the keto group at greater distances along the chain, are also shown for comparison.

PURINES, PYRIMIDINES, NUCLEOSIDES, AND NUCLEOTIDES

In this section of Table VI are shown data for some compounds important in metabolic activity and in the structure of the nucleic acids. Three principal kinds of acidic grouping are involved: (1) The ammonium groups of guanine, adenine, cytosine, and the nucleosides and nucleotides derived from them. The pK values of these are even more acidic than those of aromatic ammonium groups generally, being always lower than 5, often much lower. (2) The —OH and —NH groups of the purine and pyrimidine rings. These give pK values between 9 and 13. The value of 9.80 in adenine is probably associated with the —NH group at position 9, in the imidazole portion of the purine ring. In adenosine this position serves as the point of attachment of the ribose residue to the adenine; consequently this pK_A value is not observed in adenosine, but only a much weaker acid group (pK_A 12.5), presumably one of the hydrogens attached to the pyrimidine section of the purine ring. On account of the many possibilities for tautomerism in this ring, with the substituents found in the natural purines and pyrimidines, it is very difficult and perhaps meaningless to attempt to assign the exact configuration which is supposed to act as the proton donor corresponding to a given pK value. The matter has been further discussed, with numerous references, by Jordan (see footnote to Table VI, Section 7). (3) The primary phosphate pK_A values for the nucleotides are in the strongly acid range and have not

for the most part been accurately determined. The secondary phosphate pK_A values are almost always lower than the pK_2 of phosphoric acid (7.20), often by more than 1 pH unit. The progressively increasing values —6.05, 6.26, and 6.48—for these pK_A values in adenosine-5'-phosphate, and in the corresponding di- and triphosphates, reflect the influence of the increasing number of negative charges arising from the primary phosphate dissociations in the molecule, as the number of phosphate residues is increased. The secondary phosphate group of phosphocreatine has the unusually low pK of 4.6, owing to the influence of the neighboring positively charged methylguanidinium group.

The study of these groups is obviously of prime importance for understanding the titration curves of the nucleic acids; but this is a subject too large for us to consider here, and we refer again to the discussion by Jordan.

Effects of Variation in Dielectric Constant of Solvent on Relative Strength of Acids of Different Charge Types

The amount of electrical work involved in a proton transfer reaction must vary with the dielectric constant of the medium. The ionization of an uncharged acid, such as acetic acid or phenol, creates an anion and a cation. If it were justifiable to treat these as two charged spheres, of radii b_a and b_c , respectively, the molar increase in electrical free energy would be equal to

$$\Delta F^\circ = \frac{N\epsilon^2}{2D_0} \left(\frac{1}{b_c} + \frac{1}{b_a} \right)$$

in water of dielectric constant D_0 , according to the calculation given in Chapter 5. If we compare this with the electrical work for the same process in another medium (M) of dielectric constant D , the difference in standard free energies in the two media would be

$$(\Delta F_e^\circ)_M - (\Delta F_e^\circ)_{H_2O} = \frac{N\epsilon^2}{2} \left(\frac{1}{D} - \frac{1}{D_0} \right) \left(\frac{1}{b_c} + \frac{1}{b_a} \right) \quad (96)$$

If we take, for a rough calculation, $D_0 = 80$, and (say) $D = 18$, $b_c = b_a = 2 \times 10^{-8}$ cm, we obtain by the method of Chapter 5, p. 267, the result

$$(\Delta F_e^\circ)_M - (\Delta F_e^\circ)_{H_2O} \cong 3 \times 10^{11} \text{ erg mole}^{-1} \cong 7.2 \text{ kcal mole}^{-1}$$

This should correspond by equation (96) to an increase of pK_A by 5.3

units between two such media.¹³ Actually the values quoted for D_0 and D correspond approximately to water and to a 70% dioxane–water mixture, respectively, at 25°. The observed pK_A values for acetic acid are 4.756 in water and 8.321 in 70% dioxane, according to Harned. The difference of 3.565 in pK_A is of the same order of magnitude as that calculated from the electrical effects, using the very crude model given above. The difference could be readily adjusted by a moderate increase in the assumed radii b_c and b_a . However, if the pK_A values in different dioxane–water or alcohol–water mixtures are plotted against the reciprocal dielectric constants, the resulting curve is not linear, as this naive theory would predict, but is concave to the $1/D$ axis (Harned and Owen, 1950, p. 529). Hence, though we may conclude qualitatively that the electrical effects are important, we must also conclude that the model of charged spheres is too simple, and that nonelectrostatic effects probably also enter in. It must also be remembered that, in a mixed solvent, the molecules of the more polar component of the solvent tend to cluster around the ions, forcing away the molecules of the less polar component. Thus the composition of the solvent in the neighborhood of the ions does not correspond to that in the bulk of the solution, but varies from point to point as we move away from the surface of an ion. This does not qualitatively alter the general electrostatic picture presented here, but it would require a more complicated calculation, as we have already indicated in Chapter 5, in the discussion of Debye's theory of the salting-out effect.

The general suggestiveness of the electrostatic model is shown, by contrast, for the anilinium ($C_6H_5NH_3^+$) ion, the pK value of which is 4.62 in water at 25°, and 3.58 in 70% dioxane (see Robinson and Stokes, 1955, p. 504). Here the acid-base equilibrium involves only a transfer of a proton charge between aniline and water. No new ions are created; using the picture of spherical ions, we should expect the electrical work term to involve only the *difference* between the reciprocal radii of the $C_6H_5NH_3^+$ ion and of the H_3O^+ ion. This, of course, should give a much smaller electrostatic free energy term than for an uncharged acid forming an anion and a cation. In general, therefore, the great qualitative differences found for the effects of change of medium on the pK values of these different types of acids are of the sort that would be expected from electrostatic considerations.

Similar considerations may be extended to the amino acids. The value

¹³ It should be understood that ΔF° is so defined *in each medium* that it is equal to $-RT \ln K$, the value of K being defined as the limiting value of $(H^+)(B)/(A)$, expressed in concentration units at zero ionic strength in that medium. Such K values have been experimentally determined by Harned and others by emf measurements in different media.

of pK_1 of glycine changes from 2.350 in water to 3.965 in 70% dioxane at 25°. The reaction involved is the removal of a proton from a cation, so as to form another cation (H_3O^+) and a dipolar ion. This should involve a net increase in electrical free energy, which is primarily the free energy of the dipolar ion. This increase of free energy should become greater as the dielectric constant of the medium decreases, with a corresponding increase in pK_1 . The increase, however, should be less marked than for acetic acid or phenol, since the electrical free energy of a dipolar ion, in which the positive and negative charges are close together and partially neutralize one another, is less than that for two separate ions. One may attempt rough quantitative calculations, using for instance the "dumb-bell" model of a positively and a negatively charged sphere, with centers separated by a distance R , to represent the dipolar ion (Chapter 5, equations 24–26). All such models, however, are too simple to do more than reproduce the facts in a general way.

The value of pK_2 of a simple amino acid refers to a process in which a proton is removed from a dipolar ion, to give an organic anion and an oxonium ion. The work of separating the charges might be expected to give a positive electrostatic free energy term, and hence some moderate increase of pK as the dielectric constant decreases. The observed values of pK_2 for glycine change from 9.780 in water to 11.28 in 70% dioxane (Robinson and Stokes, 1955, p. 504). Again the change is in the expected direction.

This brief discussion can serve only as a small introduction to what has become a very broad field of study. Even for acids of the same charge type, there are differences in relative strength from one medium to another which cannot readily be interpreted along the lines we have indicated here. Moreover the acidic or basic properties of the solvent must play a major role in these phenomena, which we have not stressed here. For further discussion of acid-base equilibria in mixed and nonaqueous solvents, with numerous references, the reader may turn to Hammett (1940), Bates (1954), Chapter 6, and to less extended discussions in Harned and Owen (1950), Robinson and Stokes (1955), and Cohn and Edsall (1943), pp. 105–111.

PROBLEMS

1. Calculate the pH of a 0.1 M solution of sodium acetate, assuming pK_A (acetic acid) = 4.7 and pK_w = 14.

2. Calculate the pH of (1) a 0.1 M solution of ammonium chloride; (2) a 0.1 M solution of ammonia; (3) a solution containing 0.07 M ammonium chloride and 0.03 M ammonia; (4) a 0.1 M solution of ammonium acetate. Take pK_A = 9.3 for the ammonium acetate. Take pK_A = 9.3 for the ammonium ion and other pK values as in problem 1.

3. What is the pH of a solution containing 0.1 mole of glycine and 0.05 mole of hydrochloric acid per liter. Take pK_1 (glycine) = 2.35.

4. A solution of imidazole (Im) is titrated with hydrochloric acid. Plot the pH as abscissa against the ratio: moles of HCl /mole of $[(Im) + (ImH^+)]$. Take pK_A for $ImH^+ = 7.0$. Plot the buffer value of the solution against pH , assuming that the total concentration $C = (Im) + (ImH^+)$ remains constant at 0.1 M throughout the titration. What will the curve for the buffer value be if C is 0.05?

5. A solution is made up containing 0.03 mole of sodium benzoate (Na^+B^-) and 0.03 mole of acetic acid (HA) per liter. What is the pH of this solution at equilibrium if $pK_A(HA) = 4.7$ and $pK_A(HB) = 4.2$? What are the concentrations of HA , A^- , HB , and B^- ? In writing the equation of electrical neutrality for this system, you may consider that (H^+) and (OH^-) are both negligible.

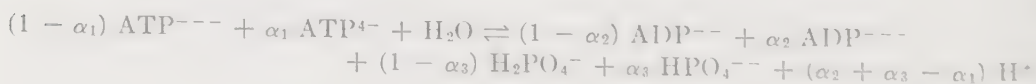
6. From the data of Table III, using equations (79) and (81) calculate pK_1 and pK_2 for glycyglycine at 0° and 38° from the values at 25° , and the values given for ΔH° and ΔC_p° . Make first a calculation assuming that ΔH° has the same value over the entire temperature range as at 25° . Then make a more exact calculation, taking account of ΔC_p° .

7. From Table III, pK_1 of β -alanine is 3.55; pK of butyric acid is 4.83 at 25° . From Bjerrum's equation (91) calculate the distance, r , from the positive charge on the ammonium group to the ionizing hydrogen of β -alanine. Westheimer and Shookhoff [*J. Am. Chem. Soc.* **61**, 555 (1939)] calculated r for β -alanine as 5.15 Å, using the Kirkwood-Westheimer theory. What "effective dielectric constant" would you have to assume to get this value of r ? (Note that Westheimer and Shookhoff assumed slightly different pK values from those given here.)

Draw a diagram of the β -alanine cation ($^+H_3N \cdot CH_2 \cdot CH_2 \cdot COOH$) in its most fully extended form, taking the $C-C$ distance as 1.53 Å, $C-N$ as 1.47 Å, $C=O$ as 1.24 Å, $C-O$ as 1.30 Å, and $O-H$ as 1 Å. Valence angles may be taken as 110° , except for the $O-C-O$ angle in the carboxyl group, which may be taken as 125° . Determine the distance from the center of the N atom, which may be taken as the center of positive charge, to the carboxyl hydrogen. This may be taken as the upper limit to the physically possible value of r .

8. Adenosine triphosphate (ATP) exists in neutral solution partly as the ion ATP^{---} , partly as ATP^{4-} . The pK value for the equilibrium $(H^+)(ATP^{4-})/(ATP^{---})$ is given in Table VI. Likewise adenosine diphosphate (ADP) exists as a mixture of ADP^{--} and ADP^{---} , with a pK value given in Table VI. Inorganic phosphate exists as a mixture of $H_2PO_4^-$ and HPO_4^{--} , with a pK value given in Table III.

The enzyme adenosine triphosphatase in muscle hydrolyzes ATP to ADP and inorganic phosphate. The reaction takes place in a buffered medium, so that the pH throughout the process may be taken as approximately constant. At a given pH , 1 mole of ATP may be regarded as consisting of $1 - \alpha_1$ mole of ATP^{---} , and α_1 moles of ATP^{4-} , where α_1 is given by equation (33), using the pK value for ATP^{---} . Likewise 1 mole of ADP consists of $1 - \alpha_2$ mole of ADP^{--} and α_2 moles of ADP^{---} ; and 1 mole of inorganic phosphate consists of $1 - \alpha_3$ mole of $H_2PO_4^-$ and α_3 moles of HPO_4^{--} , using the appropriate pK values in each case to calculate the respective values. Then the reaction which occurs when 1 mole of ATP is hydrolyzed at constant pH may be written



Calculate at pH 7.0 the quantity ($\alpha_2 + \alpha_3 - \alpha_1$) which represents the number of moles of H^+ ion, released by the above reaction, which must be neutralized by the buffers in the systems to keep the pH constant. (It may, of course, be a negative quantity.) Plot $\alpha_2 + \alpha_3 - \alpha_1$ as a function of pH between pH 6 and 7.5. [For further discussion see R. A. Alberty, R. M. Smith, and R. M. Bock, *J. Biol. Chem.* **193**, 425 (1951); and Chapter 4, pp. 210-217.]

GENERAL REFERENCES

- Bates, R. G. (1954). "Electrometric pH Determinations," John Wiley & Sons, New York. An excellent presentation of the theory and practice of pH determinations.
- Bell, R. P. (1952). "Acids and Bases," Methuen and Co., London; John Wiley & Sons, New York. A valuable short monograph on the general principles of acid-base equilibria, with an introduction to acid and basic catalysis.
- Brown, H. C., McDaniel, D. H., and Häfliger, O. (1955). Dissociation constants, in "Determination of Organic Structure by Physical Methods" (E. A. Braude and F. C. Nachod, eds.), pp. 567-662, Academic Press, New York. Contains an extraordinary range of data for organic compounds, with a correlating theoretical discussion.
- Clark, W. M. (1928). "The Determination of Hydrogen Ions," 3rd ed., Williams and Wilkins Co., Baltimore. This classic book, although thirty years old, is still valuable for reference.
- Clark, W. M. (1951). "Topics in Physical Chemistry," 2nd ed., Chapter 16, Williams and Wilkins Co., Baltimore. A useful and lucid discussion.
- Cohn, E. J., and Edsall, J. T. (1943). "Proteins, Amino Acids and Peptides," Chapters 4 and 5, Reinhold Publishing Corp., New York. Chapter 4 deals with pK values of amino acids and peptides, Chapter 5 with general relations of acidity to structure.
- Hammett, L. P. (1940). "Physical Organic Chemistry," McGraw-Hill Book Company, New York. See especially Chapters 2 and 9. An admirable discussion of fundamental principles.
- Harned, H. S., and Owen, B. B. (1950). "The Physical Chemistry of Electrolytic Solutions," 2nd ed., Chapter 15, Reinhold Publishing Corp., New York. A thorough discussion of cells without liquid junction, and tabulations of accurate data for pK values of many acids.
- Ingold, C. K. (1953). "Structure and Mechanism in Organic Chemistry," Cornell University Press, Ithaca, New York. See especially Chapter XIII. The whole book is a major contribution to its subject, with the subject of acids and bases furnishing one example of more general concepts.
- Luder, W. F., and Zuffanti, S. (1946). "The Electronic Theory of Acids and Bases," John Wiley & Sons, New York. Discussion of general theory, and illustrative data, with emphasis on concepts of G. N. Lewis.
- MacInnes, D. A. (1939). "The Principles of Electrochemistry," Reinhold Publishing Corp., New York. Contains a valuable discussion of the concept and measurement of pH .
- Pauling, L. (1940). "The Nature of the Chemical Bond," Chapter VI, Cornell University Press, Ithaca, New York. Discussion of resonance in relation to strength of acids and bases.
- Robinson, R. A., and Stokes, R. H. (1955). "Electrolyte Solutions," Academic Press, New York. See especially Chapter 12, and Appendices 12.1, 12.2, and 12.3. Similar in character to Harned and Owen (1950), but more up to date in some respects.

OTHER REFERENCES

- Bjerrum, N. (1923). *Z. physik. Chem.* **106**, 219.
Crammer, J. L., and Neuberger, A. (1943). *Biochem. J.* **37**, 302.
Greenstein, J. P. (1932). *J. Biol. Chem.* **95**, 465.
Kirkwood, J. G., and Westheimer, F. H. (1938). *J. Chem. Phys.* **6**, 506, 513.
Neuberger, A. (1937). *Proc. Roy. Soc.* **A158**, 68.
Schwarzenbach, G. (1933). *Helv. Chim. Acta* **16**, 522.
Schwarzenbach, G. (1936). *Helv. Chim. Acta* **19**, 178.
Schwarzenbach, G., and Epprecht, A. (1936). *Helv. Chim. Acta* **19**, 169.
Shugar, D. (1952). *Biochem. J.* **52**, 142.
Tanford, C., and Roberts, E. L., Jr. (1952). *J. Am. Chem. Soc.* **74**, 2509.
Westheimer, F. H., and Shookhoff, M. W. S. (1939). *J. Am. Chem. Soc.* **61**, 555.

Numerous further references are given in the tables to the present chapter; see especially Tables III and VI. For some interesting recent data, with theoretical discussion, see also E. J. King and G. W. King, *J. Am. Chem. Soc.* **78**, 1089 (1956); E. J. King, *ibid.* **78**, 6020 (1956).

Chapter 9

Polybasic Acids, Bases, and Ampholytes, Including Proteins

Dibasic Acids

Many molecules contain more than one acidic proton; some macromolecules, such as proteins and nucleic acids, may contain hundreds of such groups. They can therefore exist in a great number of intermediate ionization stages between the most acid form, in which the maximum possible number of protons is attached, and the most basic form, from

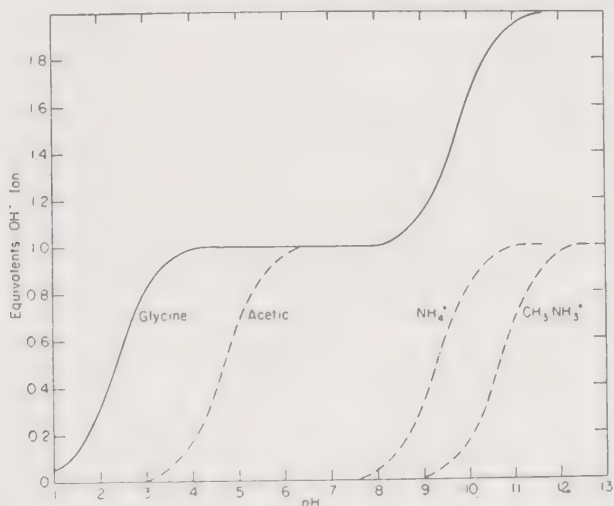


FIG. 1. Titration curves of acetic acid, methylammonium ion, and glycine.

which all acidic protons have been removed. The discussion of such polybasic acids is naturally more complicated than that of the monobasic acids which we have considered in Chapter 8. We consider first the problem of dibasic acids, which is relatively simple and may serve as an introduction to the more complicated cases which arise with polybasic macromolecules.

Consider first such a dibasic acid as the cation of a simple amino acid like glycine. This may be prepared in solution in the form of the hydrochloride of the amino acid ($+H_3N \cdot CHR \cdot COOH$)(Cl^-). The course of the titration with an added strong base at 25° is shown in Fig. 1 by the solid curve. The dotted lines, shown for comparison, represent the titra-

tion of an equimolar mixture of acetic acid and methylammonium chloride; this system resembles the amino acid in having one uncharged acid group (carboxyl) and one cationic acid group (a substituted ammonium ion). In the latter case, however, the two groups are on two separate molecules; in the former, they are both on the same molecule. Nevertheless the general form of the solid curves is strikingly similar to that of the dotted ones. Comparison would suggest, as we have assumed in Chapter 8, that the portion of the curve corresponding to pK_1 of glycine corresponds to the ionization of the carboxyl group, and pK_2 to the removal of a proton from the ammonium group.

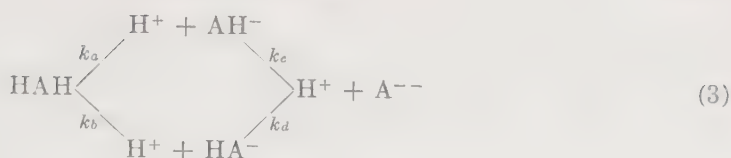
$$K_1 = \frac{(\text{H}^+)(^+\text{H}_3\text{N}\cdot\text{CHR}\cdot\text{COO}^-)}{(^+\text{H}_3\text{N}\cdot\text{CHR}\cdot\text{COOH})} = 10^{-2.35} \quad (1)$$

$$K_2 = \frac{(\text{H}^+)(\text{H}_2\text{N}\cdot\text{CHR}\cdot\text{COO}^-)}{(^+\text{H}_3\text{N}\cdot\text{CHR}\cdot\text{COO}^-)} = 10^{-9.78} \quad (2)$$

On this interpretation K_1 , for the ionization of the carboxyl group, is more than 200 times as great as the ionization constant of the carboxyl group in acetic acid ($K_A = 10^{-4.75}$). This, however, is quite reasonable on electrostatic grounds, in the light of the theoretical treatments of Bjerrum and of Kirkwood and Westheimer, already discussed in Chapter 8. All the evidence we shall consider later indeed confirms the view that equations (1) and (2) give an essentially correct description of the ionization of an amino acid such as glycine, and we shall apply them later, for instance in calculating the isoelectric point of glycine.

The situation, however, cannot be quite as simple as this. We observe that there is a finite possibility that the cation will lose a proton from the amino rather than the carboxyl group, giving the less polar molecule $\text{H}_2\text{N}\cdot\text{CHR}\cdot\text{COOH}$. Indeed in the aminobenzoic acids the dipolar ion and its less polar isomer are present in comparable amounts. A further analysis is required, if we are to obtain a full understanding of the situation.

For brevity we may denote any dibasic acid, containing two removable protons, as HAH . This may lose a proton in either one of two ways to give the two conjugate bases HA^- and AH^- , and these—which act as both acids and bases—may each lose a proton to give the same conjugate base, A^{--} . HA^- and AH^- are to be regarded as two distinct microscopic forms, corresponding to the fact that there are two distinct ways of removing a proton from HAH , or of adding one to A^{--} . They may, of course, be completely equivalent, if HAH is a symmetrical structure such as glutaric acid or the ethylenediammonium ion. The acid HAH may have any net charge, positive, negative, or zero; here its charge is written as zero for convenience. The equilibria involved, for the individual acid and basic groups, may be described by the following scheme.



The four k 's are acid dissociation constants, which involve the individual, microscopically distinct forms. It should be noted that they are not independent but are subject to the relation $k_a k_c = k_b k_d$. Hence there are only *three* independent constants involved, since if any three of the four constants are specified, the fourth is determined.

To simplify the discussion, and to concentrate on the problems which primarily concern us here, we assume for the present that all acid dissociation constants are formulated in terms of the concentrations of the components—that is, that the k 's are the "apparent constants" previously denoted by K' (equations 70 and 71 of Chapter 8). If the titration is carried out at constant ionic strength, in the presence of an excess of a neutral salt such as potassium chloride, the K 's are maintained essentially constant throughout the titration. For simplicity of notation we omit the primes and denote the various constants employed below by K , k , or G , with appropriate subscripts.

The situation is governed by two equations, one of which states the condition of electrical neutrality, the other the constancy of the total concentration (C) of the acid HAH and its conjugate bases present in all forms. The acid is converted to these more basic forms by adding the hydroxide of an alkali metal, M^+OH^- , where M^+ is generally sodium or potassium ion. The amount of M^+ present, in the equation for electrical neutrality, is equal to the concentration of M^+OH^- added to the acid solution. Then the fundamental equations are

$$(\text{HAH}) + (\text{AH}^-) + (\text{HA}^-) + (\text{A}^{--}) = C \quad (4)$$

$$(\text{M}^+) + (\text{H}^+) = (\text{AH}^-) + (\text{HA}^-) + 2(\text{A}^{--}) + (\text{OH}^-) \quad (5)$$

By making use of the mass law these two equations may be rewritten as

$$C = (\text{HAH}) \left[1 + \frac{k_a + k_b}{(\text{H}^+)} + \frac{k_a k_c}{(\text{H}^+)^2} \right] \quad (6)$$

and

$$(\text{M}^+) + (\text{H}^+) = (\text{HAH}) \left[\frac{k_a + k_b}{(\text{H}^+)} + \frac{2k_a k_c}{(\text{H}^+)^2} \right] + (\text{OH}^-) \quad (7)$$

Between them it is possible to eliminate (HAH). If we assume, as is common practice in dealing with weak acids, that (H^+) and (OH^-) are neg-

ligible, the result becomes

$$\begin{aligned} \frac{(M^+)}{C} = \bar{h} &= \frac{(AH^-) + (HA^-) + 2(A^{--})}{(HAH) + (AH^-) + (HA^-) + (A^{--})} \\ &= \frac{\frac{k_a + k_b}{(H^+)} + \frac{2k_a k_c}{(H^+)^2}}{1 + \frac{k_a + k_b}{(H^+)} + \frac{k_a k_c}{(H^+)^2}} \quad (8) \end{aligned}$$

Here we have introduced a quantity, h , to characterize each individual species in the acid-base system HAH , AH^- , HA^- , A^{--} . It is simply that number of protons which must be removed from HAH to give the species in question. Thus $h = 0$ for HAH , 1 for AH^- or HA^- , and 2 for A^{--} . The average value of h —that is, \bar{h} —which is given in equation (8), is obtained by multiplying the concentration of each species by its h value, adding up all the terms, and dividing by the total concentration of all the species. Obviously \bar{h} , which is a direct measure of the progress of the titration, must lie between zero, in strongly acid solutions, and 2, in strongly alkaline solutions.

We note that, although h for any given species in the system must be an integer (or zero), the same is not true of \bar{h} , which is a continuous function of the amount of OH^- ion added. The definitions here given for h and \bar{h} are obviously readily generalized for the case of an n -valent acid, for which h may assume any value between zero and n in the course of the titration.

Equation (8) shows how the amount of OH^- ion—in the form of $M^+ + OH^-$ —added to the system per mole of dibasic acid varies with (H^+) and with the various k 's.

Now consider a mixture of two simple monobasic acids, HA_1 and HA_2 , each at concentration C , to which strong base is added. Call the acid dissociation constants of these G_1 and G_2 , respectively. Then

$$(M^+) + (H^+) = \frac{(HA_1)G_1}{(H^+)} + \frac{(HA_2)G_2}{(H^+)} + (OH^-) \quad (9)$$

and

$$C = (HA_1) \left[1 + \frac{G_1}{(H^+)} \right] = (HA_2) \left[1 + \frac{G_2}{(H^+)} \right] \quad (10)$$

If we again neglect (H^+) and (OH^-) , these equations may be combined with elimination of (HA_1) and (HA_2) to give

$$\frac{(M^+)}{C} = \frac{\frac{G_1}{(H^+)}}{1 + \frac{G_1}{(H^+)}} + \frac{\frac{G_2}{(H^+)}}{1 + \frac{G_2}{(H^+)}} = \frac{\frac{G_1 + G_2}{(H^+)} + \frac{2G_1 G_2}{(H^+)^2}}{1 + \frac{G_1 + G_2}{(H^+)} + \frac{G_1 G_2}{(H^+)^2}} = \bar{h} \quad (11)$$

The complete parallelism between this result and the previous result obtained for the dibasic acid is apparent. In each case the individual constants appear combined in such a way as to present in reality only two over-all constants, which describe the behavior of the system. These two constants may be determined from titration data. In the case of the dibasic acid, however, these two over-all constants are composed of *three* individual constants (k_a , k_b , k_c , or k_d) and in the case of the mixture of monobasic acids of only *two*. For the mixture, therefore, the individual constants may be determined from simple titration data alone, but this is not so for the dibasic acid. In the latter case one more relation is required. This result is a special instance of a more general principle, which, without proof, may be stated as follows. The titration of any polybasic acid containing n acid groups will be indistinguishable from that of a mixture of n monobasic acids each at the same concentration as the polybasic acid, with suitably chosen dissociation constants. It will be describable in terms of n over-all constants, each of which, however, will be a function of the constants of the individual groups involved. The number of these constants will of course be greater than n . From the titration data alone, therefore, it is not possible to determine the individual constants of the polybasic acid.

Given the observed titration curve—on the assumption that the titration is carried out at constant ionic strength, so that the activity coefficients of the ions may be taken as approximately constant throughout—it is thus possible, from the equivalent form of equations (8) and (11), to choose two constants, G_1 and G_2 , in such a way that equation (11) will fit the data. These constants, G_1 and G_2 , are known as *titration constants*. Their relations to the *intrinsic constants* of the individual groups (k_a , k_b , k_c , k_d) are apparent from comparison of (8) and (11):

$$G_1 + G_2 = k_a + k_b \quad (12)$$

and

$$G_1 G_2 = k_a k_c = k_b k_d \quad (13)$$

There is also another important pair of constants for a dibasic acid, known as the *dissociation constants*, K_1 and K_2 . These are defined by the relations

$$K_1 = \frac{(\text{H}^+)[(\text{HA}^-) + (\text{AH}^-)]}{[\text{HAH}]} = k_a + k_b \quad (14)$$

$$K_2 = \frac{(\text{H}^+)(\text{A}^{--})}{[(\text{HA}^-) + (\text{AH}^-)]}; \quad \frac{1}{K_2} = \frac{1}{k_c} + \frac{1}{k_d} \quad (15)$$

and

$$K_1 K_2 = k_a k_c = k_b k_d \quad (16)$$

Thus the two K 's are related to the titration constants by the equations

$$K_1 = G_1 + G_2 \quad (17)$$

and

$$K_1 K_2 = G_1 G_2 \quad (18)$$

and in terms of K_1 and K_2 the equation for \bar{h} , the mean number of protons removed per mole of HAH, becomes from (8), (14), and (15)

$$\bar{h} = \frac{\frac{K_1}{(\text{H}^+)} + \frac{2K_1 K_2}{(\text{H}^+)^2}}{1 + \frac{K_1}{(\text{H}^+)} + \frac{K_1 K_2}{(\text{H}^+)^2}} \quad (19)$$

An important special case arises when the acid HAH is symmetrical and the two conjugate bases HA^- and AH^- thus become indistinguishable. Well-known examples are the dicarboxylic acids $\text{HOOC}(\text{CH}_2)_n\text{COOH}$ and the diammonium ions $^+\text{H}_3\text{N}(\text{CH}_2)_n\text{NH}_3^+$. (In the latter case, only compounds with $n = 2$ or greater exist.) Here it is immediately apparent from the symmetry of the structure that k_a must be equal to k_b , and k_c to k_d . Therefore in such a case the observed dissociation constants are related to the intrinsic constants of the individual groups by the relations

$$K_1 = 2k_a = 2k_b \quad (20)$$

$$K_2 = \frac{k_c}{2} = \frac{k_d}{2} \quad (21)$$

In practice it is always found that $K_1 > K_2$, and this is to be expected on electrostatic grounds. Because of the symmetry of HAH, either proton is equally likely to be given up to an added base. Once a proton has been lost from one group, however, it becomes more difficult to remove a proton from the second group. If HAH is uncharged, HA^- , because of its negative charge, tends to hold the second proton more strongly. Likewise, in a diammonium ion, the loss of one proton removes the acid-strengthening effect of the second positive charge. The larger the distance between the two acidic groups in the molecule, the smaller in general these interactions will be. Small though they may be, however, they act always in the same direction; we can never expect the removal of one proton to increase the acidity of another proton in the same molecule, and indeed such an effect is never found in practice. This is in contrast to the phenomena observed¹ in oxidation-reduction reactions, in which it is very

¹ There are a few rare cases in which the removal of a proton initiates a rather far-reaching rearrangement of molecular structure. In such cases the new structure, after rearrangement, may actually contain a stronger acid group than the original molecule. Such phenomena are rare and can generally be recognized when they occur by the

commonly found that the removal of one electron from a molecule greatly facilitates the removal of a second.

It is valuable, however, to consider the limiting case in which the two protons of a symmetrical dibasic acid are so far apart that the interaction between them vanishes. In this case $k_a = k_b = k_c = k_d$, and we may set them all equal to a single intrinsic constant, k . In this case we have for the measured constants $K_1 = 2k$ and $K_2 = k/2$; hence

$$K_1 = 4K_2 \quad (22)$$

This is the smallest possible ratio for the two successive dissociation constants for a dibasic acid. The corresponding titration constants are seen from (17) and (18) to be

$$G_1 = G_2 = \frac{K_1}{2} = 2K_2 = k \quad (23)$$

Thus the titration curve in this limiting case is identical with that of a system containing two moles of a single monovalent acid, with the same constant, k , as the intrinsic constant for each of the individual groups in the divalent acid.

The other limiting case to be considered is that in which one of the intrinsic constants, k_a and k_b , is much greater than the other. Let k_a be the greater, so that $k_a \gg k_b$, and therefore, from (3), $k_d \gg k_c$. In this case it is apparent from (17) and (18) that the titration constants and the dissociation constants become identical: $G_1 = K_1$ and $G_2 = K_2$, for all practical purposes. Here the distinction between the two kinds of constant becomes unimportant, as in the case of glycine and other amino acids already discussed.

The *form* of the titration curve of a monobasic acid, as given by Chapter 8, equation (31) or (33), is independent of the pK value; if we plot α as a function of $(pH - pK_A)$ the curves for different pK values are symmetric about the mid-point and identical with one another. Similarly for a dibasic acid, the form of the titration curve—as distinct from its

fact that the rearrangements take an appreciable period of time, in contrast to ordinary acid-base equilibria, which are practically instantaneous. For a discussion of a few such cases, see M. P. Schubert, *Ann. N. Y. Acad. Sci.* **40**, 111 (1940).

Actually the study of proteins and nucleic acids indicates that similar rearrangements—sometimes reversible, sometimes irreversible—occur within these macromolecules. Some potentially acidic groups may be virtually unreactive in the native molecule, because they are tied up in hydrogen bonds or other linkages, or are buried in the interior of the molecule where they are inaccessible to the solvent. When the structure of the native macromolecule is opened up by denaturation, these groups may become fully reactive. We shall encounter several such cases in the discussion of proteins later in this chapter. For the present, however, we confine our discussion to acid-base equilibria which are free from these complexities.

absolute position on the pH scale—is independent of the absolute values of K_1 and K_2 and depends only on their ratio. In Fig. 2 are shown plots of titration data for values of K_1/K_2 of 4, 16, 100, and 10,000. Since \bar{h} varies between 0 and 2, the mid-point of the titration comes at $\bar{h} = 1$. When $\bar{h} = 1$ is inserted into (19), it is readily seen that the mid-point corresponds to $(H^+)_{mid}^2 = K_1K_2$, or $pH_{mid} = (pK_1 + pK_2)/2$, for any value of K_1/K_2 . Thus the data for different ratios of K_1/K_2 are most

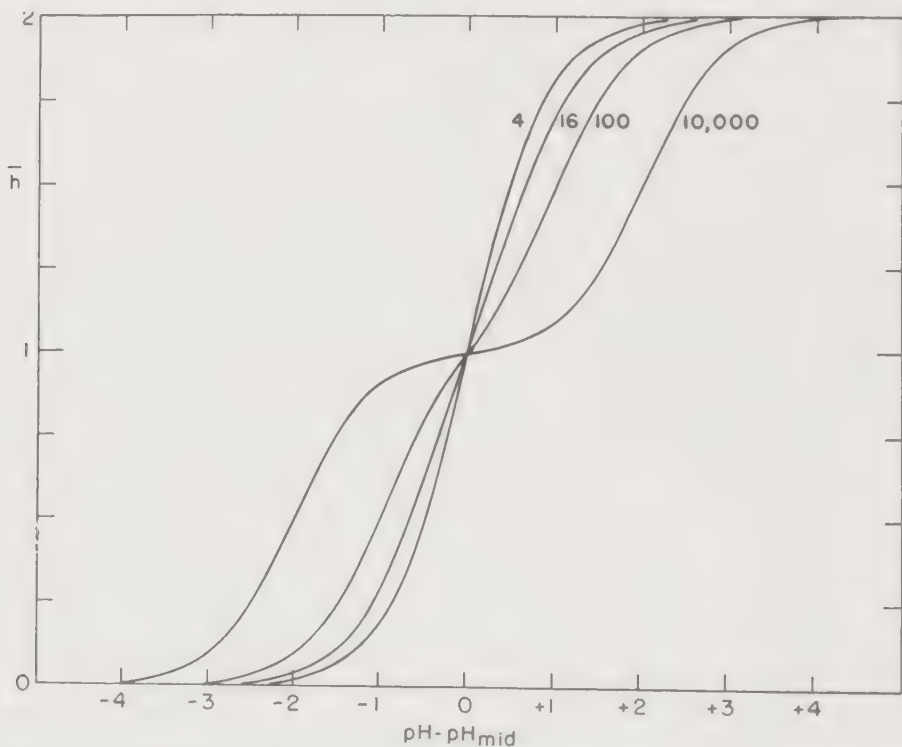


FIG. 2. Titration curves of dibasic acids, for different values of the ratio of K_1/K_2 . All curves are adjusted to coincide at the midpoint.

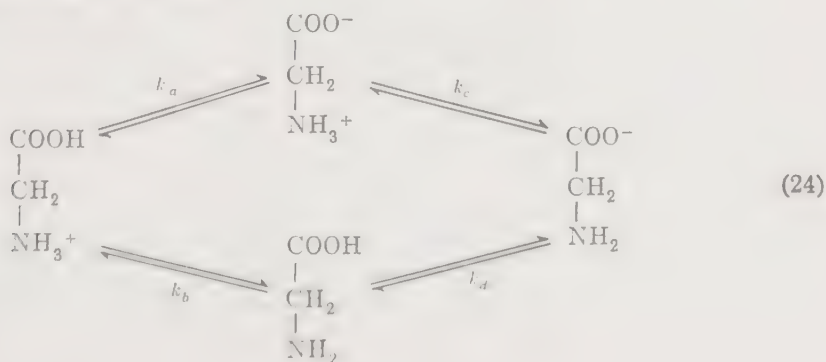
conveniently compared by plotting \bar{h} as a function of $(pH - pH_{mid})$. The limiting curve for $K_1 = 4K_2$, as already pointed out, is identical with the curve for two moles of a monobasic acid with a titration constant $G = K_1/2 = 2K_2$. The curve for $K_1 = 16K_2$ is also a smooth curve, which rises continuously in the region around the mid-point, so that good buffering action is obtained over a rather wide range of pH around the mid-point. The curve for $K_1 = 100K_2$ shows a distinct break near the mid-point, which is not present in the two curves previously described: there is strong buffering near $pH = pK_1$ and $pH = pK_2$, but the buffering near the mid-point is considerably weaker. Finally the curve for $K_1 = 10,000K_2$ shows a considerable region near the mid-point where there is virtually no buffer action, with strong buffering near $pH = pK_1$ or pK_2 .

The factor 4 in the equation $K_1 = 4K_2$, for equivalent and independent groups, deserves comment. This factor is readily derived from statistical considerations. The acid HAH can liberate a proton from either one of two positions; either of its conjugate bases, AH^- and HA^- , can take up a proton in only one. Thus there is a probability factor of 2 which favors proton liberation, if we compare HAH with a monobasic acid (HA) containing one group of equal intrinsic acid strength to the two in HAH. Likewise the base A^{--} has two receptor positions available for protons, but its conjugate acids HA^- and AH^- have only one proton to donate. Hence if k is the intrinsic constant of the acid groups involved, $K_1 = 2k$ and $K_2 = k/2$. This statistical argument is readily extended to polybasic acids.

It has been found experimentally, notably by the Austrian chemist R. Wegscheider, that the monoester $HOOC \cdot (CH_2)_n COOCH_3$ of a dicarboxylic acid, $HOOC \cdot (CH_2)_n COOH$, has in most cases a value of K_A nearly equal to half the value of K_1 for the dibasic acid. This indicates that the ester group, $-COOR$, has essentially the same effect on the acidity of a neighboring group as an undissociated $-COOH$ group, for the statistical factor just discussed is changed from 2 to 1 if one group of the acid HAH is blocked without altering the intrinsic value of K for the other. The electric moments of carboxylic acids (if they are not associated) and of their esters are indeed very nearly the same—of the order of 1.7 to 1.8 Debye units—and, as we have seen (Table IV of Chapter 8), these moments play a major part in determining the effect of a substituent group on the acidity of a neighboring group. We shall make use of this equivalence of the $-COOH$ and $-COOR$ groups, in their effect on the acidity of neighboring groups, in interpreting the ionization of the amino acids.

Equilibria between Dipolar Ions and Their Uncharged Isomers in the Amino Acids

In the case of a simple amino acid such as glycine the ionization scheme (3) assumes the form



The experimental K values obtained by titration are $K_1 = 4.45 \times 10^{-3}$ ($pK_1 = 2.35$) and $K_2 = 1.66 \times 10^{-10}$ ($pK_2 = 9.78$) at 25° (see Chapter 8, Table III; also Fig. 1 of this chapter).

To obtain the intrinsic constants, however, we must have at least one additional piece of information. This is obtainable if we assume that the constant K_E for the ammonium group in the methyl or ethyl ester of glycine is the same as the value of k_b in (24) above.

$$K_E = \frac{(H^+)(H_2N \cdot CH_2COOR)}{(^+H_3N \cdot CH_2COOR)} = \frac{(H^+)(H_2N \cdot CH_2 \cdot COOH)}{(^+H_3N \cdot CH_2 \cdot COOH)} = k_b \quad (25)$$

This application of Wegscheider's principle to the amino acids was first made by the German chemist L. Ebert (1926). The value of pK_E for glycine was found at 25° to be 7.73 for the ethyl ester and 7.66 for the methyl ester (Cohn and Edsall, 1943, Chapter 4). We take the average as 7.70; hence $K_E = 2.0 \times 10^{-8}$. We may now write, using equations (14), (15), and (16),

$$k_a = K_1 - k_b = K_1 - K_E = 4.45 \times 10^{-3} - 2 \times 10^{-8} \cong 4.45 \times 10^{-3}$$

also

$$k_c = K_2 \frac{(k_a + k_b)}{k_a} \cong K_2 = 1.66 \times 10^{-10}$$

and

$$k_d = K_2 \frac{(k_a + k_b)}{k_b} = 1.66 \times 10^{-10} \frac{(4.45 \times 10^{-3} + 2 \times 10^{-8})}{2 \times 10^{-8}} = 3.70 \times 10^{-5}$$

From this it is possible to deduce at once the ratio of the two isoelectric forms of glycine. Since at any pH each of these is in equilibrium with the acid form, $NH_3^+CH_2COOH$, they must be in equilibrium with each other:

$$\frac{(H^+)(NH_3^+CH_2COO^-)}{(NH_3^+CH_2COOH)} = k_a; \quad \frac{(H^+)(NH_2CH_2COOH)}{(NH_3^+CH_2COOH)} = k_b$$

whence the ratio of dipolar ions to uncharged molecules, K_Z , becomes

$$K_Z = \frac{k_a}{k_b} = \frac{k_d}{k_c} = \frac{(^+NH_3CH_2COO^-)}{(NH_2CH_2COOH)} = \frac{4.45 \times 10^{-3}}{2 \times 10^{-8}} = 223000 \quad (26)$$

This ratio is independent of pH . The constant K_Z , which fixes it, is the equilibrium constant for the tautomeric reaction



The large value of the constant shows that for most practical purposes we may neglect the uncharged form NH_2CH_2COOH in comparison with

the form $\text{NH}_3^+\text{CH}_2\text{COO}^-$. This, of course, is fully in accord with the studies of the dipole moments of the amino acids, and the magnitude of their electrostatic interactions with salts, which have been discussed in Chapters 5 and 6.

Polyvalent Acids: General Relations

Many macromolecules, such as proteins and nucleic acids, contain large numbers of acid and basic groups. Titrations of such polyvalent acids may be carried out, as with monovalent or divalent acids, and the results represented by plotting \bar{h} as a function of the $p\text{H}$ of the solution, taking the point of reference ($\bar{h} = 0$) as the state of the molecule when the number of bound protons is the maximum number which the molecule can bind. The simplest case to consider is the ideal limiting example of a molecule containing n acidic groups, all identical and all so far apart from one another that they ionize quite independently. Suppose that, at any instant, we could examine any particular acidic group—call it group A—on one of the molecules in a solution of a given $p\text{H}$, and determine whether it exists at that moment in the acid state ($-\text{RH}$) or as the conjugate base ($-\text{R}^-$). Let the probability be α that we should find this group to be in the latter state. The probability, α , for the same group (A) in all the other molecules in the system is, of course, the same; hence α represents the fraction of the groups of class A which are in the basic state at the given $p\text{H}$. Since by hypothesis group A ionizes quite independently of all the other groups, and all the other groups are inherently equivalent to those of class A, α as a function of $p\text{H}$ is given by the equation for a monovalent acid with the same pK value as that of group A. For reasons which will soon become apparent, we denote this value as the titration constant, pG_A :

$$p\text{H} = pG_A + \log \frac{\alpha}{1 - \alpha} \quad (27)$$

The value of \bar{h} , per mole of groups of class A, is equal to α , and thus varies from zero to unity. By our hypothesis, however, all the groups are alike; hence for an acid with n identical and independent groups the value of \bar{h} is n times as great as for any one of the groups.

$$\bar{h} = n\alpha \quad (28)$$

where α is given by (27). Thus the titration curve of such an n -valent acid is indistinguishable from that of n moles of a monovalent acid with a pK value equal to that of the titration constant, pG . The special case of a

bivalent acid with equivalent and independent groups has already been discussed (see equations 22 and 23).²

More generally, if the acidic groups are inherently different, or if the loss of a proton from one group affects the acidity of its neighbors, the curve cannot be described by a single titration constant. We may extend (27) and (28) to another idealized case, in which we assume that there are n_A independent and identical groups of class A, with titration constant G_A ; n_B such groups of class B, with titration constant G_B , etc. In this case there is an equation of the form of (27) for each class of group, with a corresponding value of α ($\alpha_A, \alpha_B, \dots$) defined for any given pH . The value of \bar{h} then becomes (see 28 above)

$$\bar{h} = n_A \alpha_A + n_B \alpha_B + \dots = \sum_i n_i \alpha_i \quad (29)$$

Suppose, for instance, that a hypothetical peptide contains n_A carboxyl groups (of aspartic or glutamic residues), all with $pG = 4$, and n_B ammonium groups (for instance, ϵ -ammonium groups of lysine residues), all with $pG = 10$. Here it is plain that, at pH 7, to a close approximation $\alpha_A = 1$ and $\alpha_B = 0$. In this case, where the two pG values are widely separated, it is not necessary in applying (29) to assume that the carboxyl and ammonium groups are independent of one another. They might be placed in pairs, each $-\text{COOH}$ closely adjoining an $-\text{NH}_3^+$, so that each strongly influenced the ionization of the other. Equation (29) would still apply, provided that all the $-\text{COOH}$ groups ionize independently of one another, at pH values well below those at which the $-\text{NH}_3^+$ groups start to lose their protons; and similarly that all the NH_3^+ groups yield up their protons independently of one another, in the presence of the completely ionized carboxyls. Any such model is, of course, highly idealized; the electrostatic forces produced by the addition or removal of protons from the molecule have a significant influence on the tendency to bind or release other protons at other combining sites. We shall consider the calculation of such effects later in this chapter. When the number and chemical nature of the various kinds of acidic groups in a large molecule is known, however, equation (29) is often useful in making a rough estimate of the form of the titration curve.

Regardless of such special assumptions, however, we may formulate the titration constants ($G_1, G_2 \dots G_n$), the macroscopic dissociation constants ($K_1, K_2 \dots, K_n$), and the intrinsic constants ($k_1, k_2 \dots k_n$) of individual groups of an n -valent acid, containing different groups of

² The proposition stated in (27) and (28) has been proved mathematically in various ways—see Simms (1926); von Muralt (1930); Wyman in Cohn and Edsall (1943, pp. 451–453). A valuable general discussion is given by Klotz (1953).

any sort we choose, by a natural extension of the definitions given already for the special case in which $n = 2$. If there are n acidic groups in the molecule, considered in its most acid state ($h = 0$) with the maximum number of protons attached, then there are n different ways in which a proton may be removed from this acid. We denote the most acid form of the molecule by the symbol 0A , and each of the n conjugate bases, from which one proton has been removed, by the symbol ${}^1A_1, {}^1A_2 \dots {}^1A_n$. (The superscript at the left denotes the value of h for the species in question; the subscript at the right indicates the group from which a proton has been removed.) The equilibrium between 0A and the set ${}^1A_1, {}^1A_2$, etc., involves n microscopic constants, which may be written

$$k_1 = \frac{(H^+)({}^1A_1)}{({}^0A)}; \quad k_2 = \frac{(H^+)({}^1A_2)}{({}^0A)}; \text{ etc.} \quad (30)$$

The first macroscopic dissociation constant, K_1 , involves the equilibrium between 0A and *all* molecules from which a single proton has been removed ($h = 1$). We denote this sum by

$$\begin{aligned} (\Sigma {}^1A) &= ({}^1A_1) + ({}^1A_2) + \dots + ({}^1A_n) \\ K_1 &= \frac{(H^+) \left(\sum_{i=1}^n {}^1A_i \right)}{({}^0A)} = k_1 + k_2 + \dots + k_n \end{aligned} \quad (31)$$

Consider now the set of microscopically different species, each of which contains 2 less protons than 0A . The two groups from which protons have been lost, in one particular species, are groups i and j ; we denote this species of molecule by the symbol ${}^2A_{ij}$. Here again the superscript denotes the value of h . The equilibria between the various species 1A_i and the species ${}^2A_{ij}$ can be formulated in terms of a set of microscopic constants:

$$k_{12} = \frac{(H^+)({}^2A_{12})}{({}^1A_1)}; \quad k_{21} = \frac{(H^+)({}^2A_{12})}{({}^1A_2)}; \quad k_{34} = \frac{(H^+)({}^2A_{34})}{({}^1A_3)}; \text{ etc.} \quad (32)$$

Here, for instance, k_{34} denotes the equilibrium constant for a process in which a proton is removed from group 4, in a molecule which has already lost a proton from group 3. The number of microscopically distinct species of the class ${}^2A_{ij}$ is given by the number of ways in which two objects can be selected, without regard to order of selection, from a set of n distinguishable objects. This is equal to $n!/[2!(n-2)!] = n(n-1)/2$. The number of microscopic constants with two subscripts, k_{ij} , is equal to $n(n-1)$, since there are n kinds of molecules of the class 1A_i , and a proton can be removed from each of these in $n-1$ different ways to give a

molecule of the class ${}^2A_{ij}$. There are obvious interrelations between the microscopic constants; for instance, using equations (30) and (32), it is readily seen that $k_1k_{12} = k_2k_{21}$, and in general $k_ik_{ij} = k_jk_{ji}$. We denote the sum of the concentrations of all molecules of the class ${}^2A_{ij}$ by the symbol $(\Sigma {}^2A)$.

The second macroscopic dissociation constant, K_2 , is given by the relation

$$K_2 = \frac{(H^+)(\Sigma {}^2A)}{(\Sigma {}^1A)} \quad (33)$$

where the summation in the numerator includes all the different microscopic species for which $h = 2$, and that in the denominator includes all the different microscopic species for which $h = 1$. Any of the n macroscopic constants—for instance K_q —is likewise given by a similar formula:

$$K_q = \frac{(H^+)(\Sigma {}^qA)}{(\Sigma {}^{(q-1)}A)} \quad (34)$$

From (31), (33), and (34) we may also write a relation which will shortly be found useful:

$$K_1K_2 \cdots K_q = \frac{(H^+)^q(\Sigma {}^qA)}{({}^0A)} \quad (35)$$

The numerator of (34) contains the sum of the concentrations of *all* microscopic species for which $h = q$, the denominator the corresponding sum for all the species for which $h = q - 1$. The number of different species for which $h = q$ is $n!/[(n - q)!q!]$.

The microscopic constants corresponding to K_q are of the form

$$k_{12\dots q} = \frac{(H^+)({}^qA_{12\dots q})}{({}^{(q-1)}A_{12\dots q-1})} \quad (36)$$

There are $n!/[(q - 1)!(n - q + 1)!]$ microscopic species of the class $h = q - 1$. Each of these still contains $n - q + 1$ dissociable protons, and any one of these may be removed to give one of the species $h = q$. Therefore the total number of microscopic constants corresponding to K_q is $n!/[(q - 1)!(n - q)!]$. Any given microscopic constant is actually specified (1) by the last number in the subscript, which denotes the position of the proton involved in the reaction under consideration, and (2) by the whole set of preceding numbers in the subscript, which—regardless of order—denote the other positions from which protons have already been removed. Thus k_{1346} is the equilibrium constant for removal of a proton from position 6, in a molecule from which the protons at positions 1, 3, and 4 have already been removed. Thus k_{1346} , k_{1436} , k_{3146} , k_{4136} , and

k_{4316} are all identical. It is generally most convenient, in such cases, to write the numbers in the subscript in ascending order, except of course for the number referring to the proton actually being removed in the process under consideration, which must always be placed last.

In all classes ($h = 0$ to $h = n$) there are 2^n microscopically distinct species for an n -valent acid, and $n \cdot 2^{n-1}$ microscopic k values.

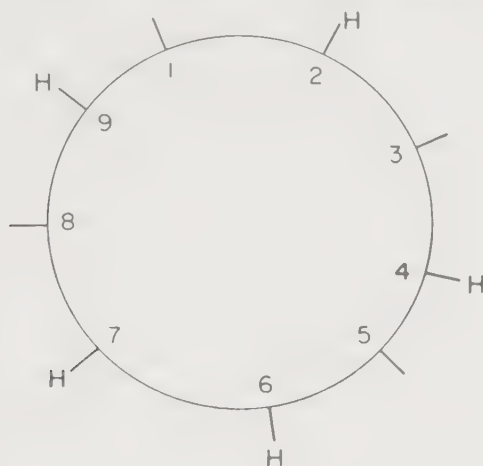


FIG. 3. A particular microscopic form of a molecule with nine basic sites, of which five are occupied by protons.

We may note one of the relations between the products of the microscopic and macroscopic constants:

$$k_1 k_{12} \cdots k_{12\dots n} = K_1 K_2 \cdots K_n = \frac{(\text{H}^+)^n ({}^n A)}{({}^0 A)} \quad (37)$$

It may be useful to illustrate these general relations by a specific example. Consider a molecule (Fig. 3), containing nine basic sites capable of binding protons: these are numbered from 1 to 9. In the particular microscopic form shown, sites 2, 4, 6, 7, and 9 are occupied by protons; sites 1, 3, 5, and 8 are unoccupied. The value of h for this molecular species is therefore 4; it may be denoted by the symbol ${}^4 A_{1358}$. There are altogether $9!/5!4!$ or 126 different microscopic species of the class ${}^4 A$. Each of these contains 5 dissociable protons, so that there are $5 \times 126 = 630$ different microscopic dissociation constants involving the removal of a proton from a molecule of the class ${}^4 A$ to give a conjugate base of the class ${}^5 A$. Thus the species obtained by removing the proton at position 4 from the species represented in Fig. 3 would be denoted as ${}^5 A_{13584}$, or ${}^5 A_{13458}$, and the corresponding microscopic equilibrium constant is

$$k_{13584} = \frac{(\text{H}^+) ({}^5 A_{13584})}{({}^4 A_{1358})}$$

The relatively simple cases of two amino acids, each containing 3 dissociating protons, are discussed in detail below. The system of notation which we are employing here was first suggested by T. L. Hill (see references in Table II below).

The quantity determined in a titration curve, when pH is measured as a function of hydrogen or hydroxyl ions added, can be expressed as the mean value, \bar{h} , of the quantity h , per mole of polyvalent acid present, averaged over all molecules in the system which is being titrated. The contribution to \bar{h} from all ions of the class from which q protons have been removed is $q(\Sigma {}^qA)$. We can then write \bar{h} in terms of the K 's by summing the contributions of all the ions of all h values and dividing by the total concentration of the polyvalent acid in all its forms from 0A to nA .

$$\begin{aligned}\bar{h} &= \frac{(\Sigma {}^1A) + 2(\Sigma {}^2A) + \cdots + n({}^nA)}{({}^0A) + (\Sigma {}^1A) + (\Sigma {}^2A) + \cdots + (\Sigma {}^{(n-1)}A) + ({}^nA)} \quad (38) \\ &= \frac{\frac{K_1}{(H^+)} + \frac{2K_1K_2}{(H^+)^2} + \frac{3K_1K_2K_3}{(H^+)^3} + \cdots + \frac{nK_1K_2 \cdots K_n}{(H^+)^n}}{1 + \frac{K_1}{(H^+)} + \frac{K_1K_2}{(H^+)^2} + \cdots + \frac{K_1K_2 \cdots K_n}{(H^+)^n}} \\ &= \frac{\sum_{h=0}^n \frac{hL_h}{(H^+)^h}}{\sum_{h=0}^n \frac{L_h}{(H^+)^h}} = \frac{\partial \log_{10} 1}{\partial \log \frac{1}{(H^+)}} \sum_{h=0}^n \left[\frac{L_h}{(H^+)^h} \right] = \frac{\partial \log_{10}}{\partial pH} \sum \frac{L_h}{(H^+)^h}\end{aligned}$$

In the summation given by the last expression of equation (38) we have for brevity introduced the constants $L_0 = 1$, $L_1 = K_1$, $L_2 = K_1K_2$, etc. We note that no summation sign is written for the molecules of the class nA in the first formula for \bar{h} in (38), since as with 0A there is only one microscopic form corresponding to this class.

The titration constants G_1, G_2, \dots, G_n are related to the macroscopic dissociation constants by the following relations:

$$\begin{aligned}K_1 &= G_1 + G_2 + \cdots + G_n = \sum G_i \\ K_1K_2 &= G_1G_2 + G_1G_3 + \cdots + G_2G_3 + \cdots = \sum_{i,k} G_iG_k \quad (i \neq k) \\ &\quad \dots \\ K_1K_2 \cdots K_n &= G_1G_2 \cdots G_n\end{aligned} \quad (39)$$

It is also possible, and sometimes useful, to express all the equations of (39) in terms of a single formula, using the definitions $L_0 = 1$, $L_1 = K_1$, $L_2 = K_1K_2$, etc., employed in (38):

$$\sum_{i=0}^n \frac{L_i}{(H^+)^i} = \left[1 + \frac{G_1}{(H^+)}\right] \left[1 + \frac{G_2}{(H^+)}\right] \cdots \left[1 + \frac{G_n}{(H^+)}\right] \quad (39.1)$$

If the product on the right-hand side of this equation is expanded, and the coefficient of each power of (H^+) on the left-hand side of the equation set equal to the corresponding coefficient on the right, the resulting equations are exactly the equations of (39).

On combining the last expression in (38) with (39.1) we obtain a relation expressing \bar{h} as a function of (H^+) and the titration constants:

$$\bar{h} = \sum_{i=1}^n \frac{\partial \log [1 + G_i/(H^+)]}{\partial pH} \quad (39.2)$$

The proof of (39) and (39.1) has been given by Simms and by von Muralt (see footnote 2), and we shall not reproduce it here. It is worth noting, however, that these equations assume a particularly simple form in the special case of n equivalent and independent groups. In this case, as we have already seen (equation 27), there is only a single titration constant, G ; that is, $G_1 = G_2 = \cdots = G_{n-1} = G_n$. Then (39) becomes

$$\begin{aligned} K_1 &= nG \\ K_2 &= \frac{n-1}{2} G \\ K_3 &= \frac{n-2}{3} G \\ &\vdots \\ K_h &= \frac{n-h+1}{h} G \\ &\vdots \\ K_n &= \frac{G}{n} \\ K_1 K_2 \cdots K_h &= \frac{n!}{(n-h)!h!} G^h \end{aligned} \quad (40)$$

Equation (40) can also be derived, by a more illuminating procedure, from the general equations for the microscopic dissociation constants for a polyvalent acid. If all the groups are equivalent and independent, this means that all the individual microscopic dissociation constants are identical, and equal to the titration constant, G . Equation (31) then becomes

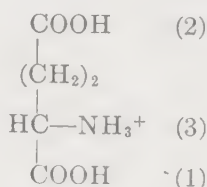
immediately $K_1 = nG$. Next consider one of the microscopic constants of the class given in equation (32), such as $k_{12} = (H^+)(^2A_{12})/(^1A_1)$. This is also equal to G , as are k_{21} , k_{34} , etc. There are $n!/2!(n-2)! = n(n-1)/2$ different microscopic species of ions of the class 2A . Since all the intrinsic constants are identical, the concentrations of all these microscopic species are the same, and the total concentration ($\Sigma ^2A$) is equal to $(^2A_{12})n(n-1)/2$. The total concentration of ions of the class 1A is, by the same reasoning, equal to $n!/(n-1)! = n$ times the concentration of any one microscopic form of this class, such as 1A_1 . Thus we can write for the second macroscopic constant, K_2 , of equation (33), in terms of a microscopic constant such as k_{12} ,

$$K_2 = \frac{(H^+)(\Sigma ^2A)}{(\Sigma ^1A)} = \frac{(H^+)(^2A_{12})}{(^1A_1)} \frac{n(n-1)}{2 \cdot n} = G \left(\frac{n-1}{2} \right) \quad (40.1)$$

This is identical with the expression given in (40). The same argument can be extended to all the other equations given in (40), if we remember that the total concentration ($\Sigma ^hA$) of all the ions of class h (where $0 \leq h \leq n$) is $n!/[h!(n-h)!]$ times as great as the concentration of any one microscopic species of the class hA .

A Specific Case: Glutamic Acid

To illustrate the general relations between microscopic and macroscopic constants, given in the preceding section, we consider the case of glutamic acid ($n = 3$), for which it is possible to estimate all the microscopic constants, due to the work of Neuberger (1936). Glutamic acid, when it contains the maximum number of bound protons, has the formula



Of the three acidic groups, we denote the α -carboxyl as 1, the γ -carboxyl as 2, and the ammonium group as 3. The acid or basic state of each group may be denoted by a charge symbol, which is 0 or $-$ for an uncharged acid group such as 1 or 2, and $+$ or 0 for a cationic acid group such as 3. We write the charge symbol for groups 1, 2, and 3 in order from left to right, and enclose the set of symbols in parentheses to denote the concentration of the species in question. Thus the concentration of the most acid species, in the formula above, may be abbreviated to (0 0 +). The scheme for the ionization of all the different microscopic forms may then be represented as shown in Table II.

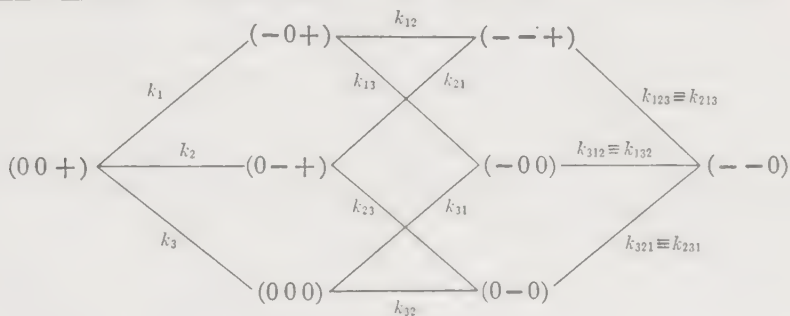
TABLE I
 pK VALUES FOR GLUTAMIC ACID AND ITS ESTERS AT 25°

Substance	pK_1	pK_2	pK_3
Glutamic acid	2.155	4.324	9.960
α -Ethyl hydrogen glutamate	3.846	7.838	
γ -Ethyl hydrogen glutamate	2.148	9.19	
Diethylglutamate	7.035		

These values are extrapolated to zero ionic strength. Data from Neuberger; see Table II below.

TABLE II
 IONIZATION SCHEME FOR GLUTAMIC ACID

Class symbol (h value)	0A	1A	2A	3A
h (protons dissociated)	0	1	2	3
z (net charge)	+1	0	-1	-2



$$k_1 = \frac{(H^+)(-O+)}{(OO+)}, \text{ etc.}; k_{12} = \frac{(H^+)(- - +)}{(-O+)}; k_1 k_{12} = \frac{(H^+)^2(- - +)}{(OO+)} = k_2 k_{21}, \text{ etc.}$$

$$K_1 = k_1 + k_2 + k_3$$

$$K_1 K_2 = k_1 k_{12} + k_1 k_{13} + k_2 k_{23} \equiv k_2 k_{21} + k_3 k_{31} + k_3 k_{32}$$

$$K_1 K_2 K_3 = k_1 k_{12} k_{123} = \frac{(H^+)^3(- - O)}{(OO+)} = k_2 k_{23} k_{231} = k_3 k_{31} k_{312}$$

Macroscopic Dissociation Constants

$$K_1 = \frac{(H^+)[(-O+) + (O-+) + (OOO)]}{(OO+)}$$

$$K_2 = \frac{(H^+)[(- - +) + (-OO) + (O - O)]}{[(-O+) + (O-+) + (OOO)]}$$

$$K_3 = \frac{(H^+)(- - O)}{[(- - +) + (-OO) + (O - O)]}$$

[See A. Neuberger, *Biochem J.* **30**, 2085 (1936); T. L. Hill, *J. Phys. Chem.* **48**, 101 (1944); *J. Chem. Phys.* **12**, 56, 147 (1944).]

In this particular case, sufficient data are available to evaluate all the microscopic constants, provided we make the same assumption already made in the case of glycine and the simple monoamino monocarboxylic acids—namely that an ester ($-\text{COOR}$) group has the same effect as a $-\text{COOH}$ group on the ionization of a neighboring group. Careful measurements by Neuberger at 25° have given the pK values for glutamic acid and its three ethyl esters recorded in Table I.

From these experimental values, and from the ionization scheme given in Table II, all the microscopic constants can be calculated; the results obtained are given in Table III.

TABLE III
MICROSCOPIC pK VALUES FOR GLUTAMIC ACID AT 25°

Ionizing group			
$\alpha\text{-COOH}$	$\gamma\text{-COOH}$	$-\text{NH}_3^+$	
$pk_1 = 2.15$	$pk_2 = 3.85$	$pk_3 = 7.04$	COOH (2)
$pk_{21} = 2.62$	$pk_{12} = 4.32$	$pk_{13} = 9.19$	$ $ $(\text{CH}_2)_2$
$pk_{31} = 4.30$	$pk_{32} = 4.65$	$pk_{23} = 7.84$	$ $ $\text{HC}-\text{NH}_3^+ \text{ (3)}$
$pk_{231} = 4.74$	$pk_{132} = 5.09$	$pk_{123} = 9.96$	$ $ COOH (1)

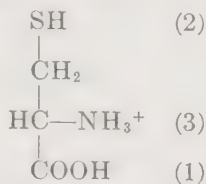
Data rearranged from T. L. Hill (see references in Table II).

To understand the significance of these values clearly, the reader should remember that the ionizing group, for each microscopic pK value, is indicated by the *last* number in the subscript. Any number preceding this in the subscript denotes another specified group in the molecule which is already in the basic form, when the ionization under consideration is taking place. Thus pk_{32} denotes the pK value of the γ -carboxyl group (group 2) in glutamic acid, when the $-\text{NH}_3^+$ group (group 3) has already been converted to the conjugate base $-\text{NH}_2$. Since the number 1 does not appear in the subscript 32, its absence denotes that group 1, the α -carboxyl, is still in the un-ionized form during the reaction corresponding to the pK value in question. Numerous relations concerning the effect of neighboring substituent groups on the ionization of a particular group can be discerned by careful examination of the data in Table III.

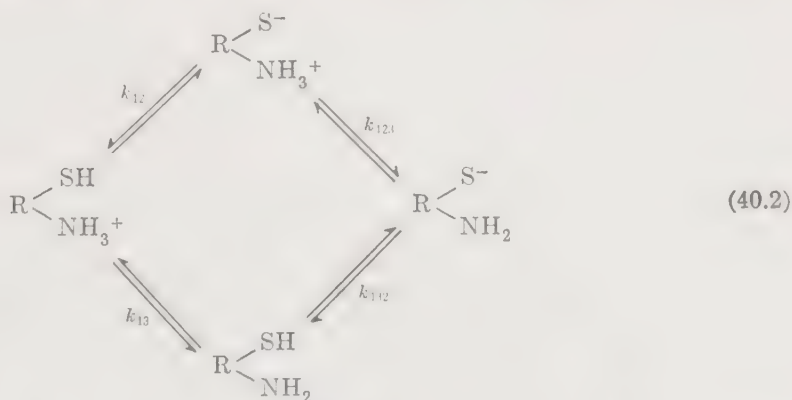
Spectroscopic Determination of Microscopic Constants: The Ionization of Cysteine

In some cases the determination of individual microscopic ionization constants can be carried out by making use of the changes in absorption spectrum which may occur on removal of a proton from one of the particular acidic groups in a system. (See Chapter 8, p. 426.) In favorable

cases these may be compared with the more indirect calculations which we have already illustrated. A particularly successful example of the application of such measurements is cysteine, the structure of which in strongly acidic media may be written



Cysteine therefore contains three acidic groups, the carboxyl, the sulfhydryl, and the ammonium, which we denote arbitrarily by the numbers 1, 2, and 3, respectively. Correspondingly the pH titration shows three pK values which have been reported by Borsook, Ellis, and Huffman (see Chapter 8, Table VI, Part 3) as 1.71, 8.33, and 10.78 at 25° . These are extrapolated values at zero ionic strength. The experimental measurements involved were simply pH determinations as a function of the amount of strong acid or alkali added to the isoelectric cysteine, with corrections when necessary for free H^+ or OH^- ion which had not been neutralized by the cysteine. It can be taken as certain from all the evidence given in this and the preceding chapter that pK_1 represents essentially the ionization of the carboxyl group. The relations of pK_2 and pK_3 to the $-\text{SH}$ and the $-\text{NH}_3^+$ groups are not immediately apparent, since the acid strength of the ammonium group is of the same order of magnitude as that of the sulfhydryl. In any case, however, the ionization scheme involved can be written in a form exactly analogous to that already given for glycine:



$$K_2 = k_{12} + k_{13} \quad \text{and} \quad K_2 K_3 = k_{12} k_{123} \equiv k_{13} k_{132} \quad (40.3)$$

The residue in this particular case, R , includes the ionized carboxyl group, which is present in all the microscopic forms under consideration. In this system both direct and indirect methods have been used to

evaluate the ratio

$$\left(\begin{array}{c} \text{S}^- \\ \diagup \quad \diagdown \\ \text{R} \\ \diagdown \quad \diagup \\ \text{NH}_3^+ \end{array} \right) / \left(\begin{array}{c} \text{SH} \\ \diagup \quad \diagdown \\ \text{R} \\ \diagdown \quad \diagup \\ \text{NH}_2 \end{array} \right) = \frac{k_{12}}{k_{13}} = \frac{k_{132}}{k_{123}}$$

For example, Rykhan and Schmidt (1944) made use of the assumption that the effect of an S-alkyl group on the ionization of a neighboring group should be approximately the same as that of an un-ionized SH group.³ Rykhan and Schmidt determined the pK for the ammonium group in S-ethylcysteine as 8.60 at 25°. We therefore take this as equal to the microscopic constant pk_{13} for cysteine itself. With the values for the macroscopic constants (pK_2 and pK_3) listed above, this led to the following constants for cysteine: $pk_{12} = 8.66$, $pk_{123} = 10.45$, $pk_{132} = 10.51$.

Grafius and Neilands (1955) examined cysteine betaine which contains a $-\text{N}(\text{CH}_3)_3^+$ group instead of the $-\text{NH}_3^+$ group of cysteine. The trimethylammonium group of course does not release a proton in the pH range below 14, but it might be expected to resemble the $-\text{NH}_3^+$ group in its effect on the ionization of a neighboring $-\text{SH}$ group. The following pK' values of cysteine, cysteine betaine, and S-methylcysteine were found:

TABLE IV
 pK' VALUE FOR CYSTEINE AND SOME OF ITS DERIVATIVES

Substance	pK'
Cysteine, pK_2'	8.30
Cysteine, pK_3'	10.40
Cysteine betaine ($-\text{SH}$)	8.65
S-Methylcysteine ($-\text{NH}_3^+$)	8.75

Note that these pK' values were determined at 0.15 ionic strength. Hence they differ appreciably from extrapolated values estimated at zero ionic strength. From Grafius and Neilands (1955).

³ A strong piece of supporting evidence for this assumption is the near identity of the pK value of the carboxyl group in thioglycolic acid, $\text{HS}\cdot\text{CH}_2\text{COOH}$ (3.67), and S-methylthioglycolic acid, $\text{H}_3\text{CS}\cdot\text{CH}_2\text{COOH}$ (3.72), as pointed out by Brown *et al.* (1955). These authors have also noted in contrast the interesting point that the pK value of glycolic acid, $\text{HO}\cdot\text{CH}_2\text{COOH}$ (3.83), is 0.3 pH unit higher than that of its methyl ether, $\text{H}_3\text{CO}\cdot\text{CH}_2\text{COOH}$ (3.53). Perhaps this is due to the increased hydrogen bonding with the solvent around the OH group, as compared with the more restricted type of bonding which is possible with the $-\text{OCH}_3$ group.

Data of high precision, on the ionization of the O-methyl derivatives of fatty acids and amino acids, were reported by E. J. King at the meeting of the American Chemical Society in New York, September 1957. These essentially confirm the statements given above, and provide new and extensive thermodynamic data.

If the pK value of cysteine betaine is taken as equal to pk_{12} , and that of S-methylcysteine as pk_{13} , it is of course possible to derive immediately the values of pk_{123} and pk_{132} . Moreover, the relation $K_2 = k_{12} + k_{13}$ is found to hold within 0.1 pH unit. These data indicate that the pK values experimentally measured for cysteine are hybrid constants. In other words the pk of the SH group in the presence of a neighboring $—NH_3^+$ group (denoted by pk_{12}) is about 2 pH units more acid than in the ionization step denoted by pk_{132} in which the neighboring amino group is uncharged. Likewise the pk of the NH_3^+ group is about 2 pH units more

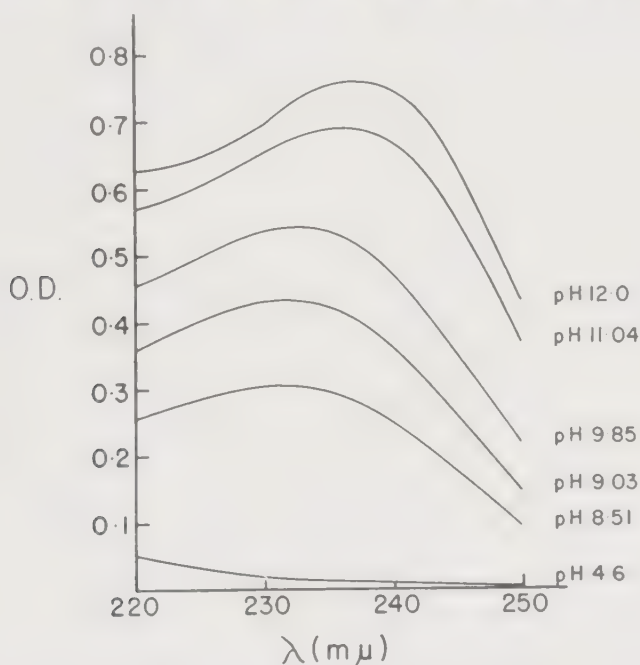


FIG. 4. The absorption spectra of cysteine solutions in the ultraviolet, as a function of pH . (From Benesch and Benesch, 1955.)

acid in the presence of a neighboring SH group than it is in the presence of the negatively charged S^- group. These relations appear reasonable in the light of the general correlations between structure and pK values which have been outlined in Chapter 8.

A more direct approach to the problem was made by Benesch and Benesch (1955), making use of the ultraviolet absorption spectrum of the ionized SH group. The nature of the effect observed is shown in Fig. 4, which shows the optical density of cysteine solutions between 220 and 250 $m\mu$ at various pH values. From this, as from absorption spectra of other compounds containing sulfhydryl groups, it is apparent that the $—S^-$ group absorbs in the ultraviolet with a maximum in the range between 232 and 238 $m\mu$. The height of this maximum, as measured by the

molar extinction coefficient, is very nearly independent of pH , with a value of approximately 5000. The position of the maximum shifts from about 231 $m\mu$ at relatively low pH to 237 $m\mu$ at high pH . This change is due apparently to the change in the state of ionization of the amino group in the cysteine molecule. This is indicated by the fact that the position of the maximum is quite independent of pH in the case of thioglycolic acid which has no adjoining amino group. From data such as those of Fig. 4, the fraction of the cysteine sulfhydryl groups in the form of RS^- —which we denote by α_{SH} —could be determined at each pH from the ratio of the absorption at the given pH value to the absorption when the SH group is completely dissociated, as it is in the pH range between 12 and 13.

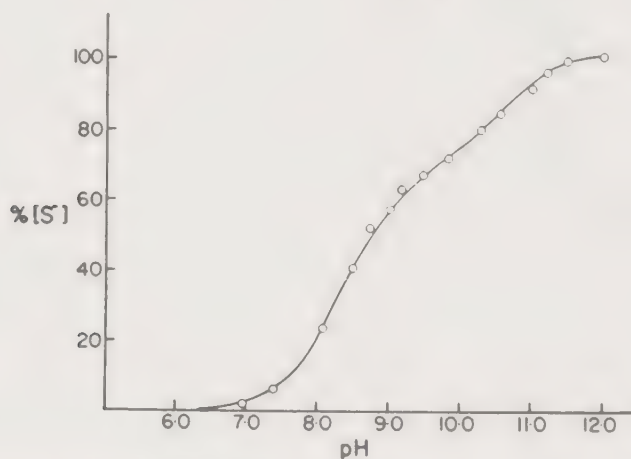


FIG. 5. Degree of ionization of the sulfhydryl group of cysteine (α_{SH}) as a function of pH . (From Benesch and Benesch, 1955.)

The result is shown in Fig. 5. The value of α_{SH} is related to the microscopic and macroscopic ionization constants by the equations

$$\begin{aligned}
 \alpha_{SH} &= \frac{(RS^-)}{(RS^-) + (RSH)} \\
 &= \frac{(S^- - R - NH_3^+) + (-S - R - NH_2)}{(HS - R - NH_3^+) + (-S - R - NH_3^+) + (HS - R - NH_2) + (-S - R - NH_2)} \\
 &= \frac{\frac{k_{12}}{(H^+)} + \frac{k_{12}k_{123}}{(H^+)^2}}{1 + \frac{k_{12} + k_{13}}{(H^+)} + \frac{k_{12}k_{123}}{(H^+)^2}} = \frac{\frac{k_{12}}{(H^+)} + \frac{K_2K_3}{(H^+)^2}}{1 + \frac{K_2}{(H^+)} + \frac{K_2K_3}{(H^+)^2}} \quad (40.4)
 \end{aligned}$$

By making use of the relation $K_2K_3 = k_{12}k_{123} = k_{13}k_{132}$, between the microscopic and macroscopic constants, it is possible, as shown above,

to eliminate all but one of the microscopic constants from the equation for α_{SH} . Rearranging (40.4), we then obtain k_{12} as a function of α_{SH} , (H^+) , K_2 , and K_3 —all of which are quantities accessible to experiment:

$$k_{12} = \alpha_{\text{SH}}[(\text{H}^+) + K_2] - (1 - \alpha_{\text{SH}}) \frac{K_2 K_3}{(\text{H}^+)} \quad (40.5)$$

Once K_1 and K_2 are known from an independent titration, each value of (H^+) , with the associated value of α_{SH} from ultraviolet absorption, fixes a value for k_{12} from (40.5). A series of such values may be obtained and suitably averaged to obtain a most probable value for k_{12} . From this, the other microscopic constants are readily determined from (40.3). The

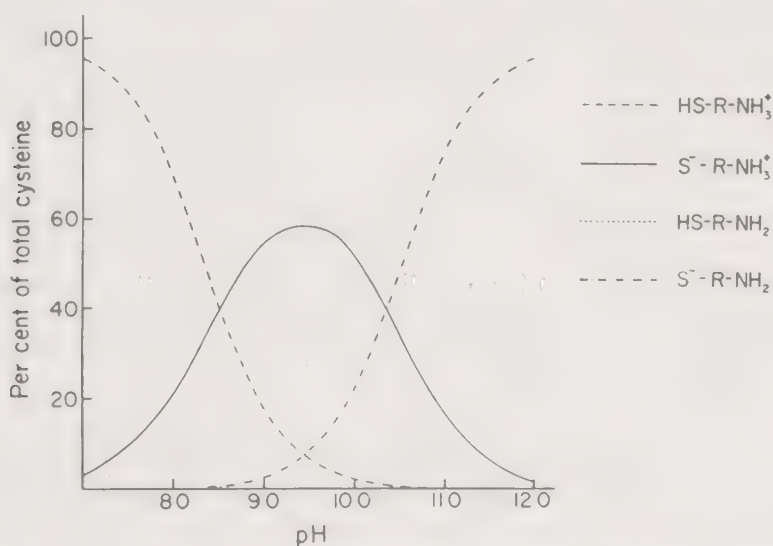


FIG. 6. Relative concentrations of different microscopic forms of cysteine at different $p\text{H}$ values. The carboxyl group is ionized in all these forms. (From Benesch and Benesch, 1955.)

values derived are shown in Table V, and from these the relative concentrations of each of the microscopic forms can be calculated as a function of $p\text{H}$.⁴ The results are shown in Fig. 6.

It will be seen that this direct method of estimating the microscopic constants gives values which differ somewhat from those obtained by the more indirect reasoning involved in the calculations of Rykman and Schmidt or of Grafius and Neilands. According to the results of Benesch and Benesch the concentration of the form $^-\text{S}-\text{R}-\text{NH}_3^+$ at any given $p\text{H}$ is about twice as great as that of the isomeric form $\text{HS}-\text{R}-\text{NH}_2$. The indirect calculation, from the data of Table IV, indicated that the

⁴ The method of deriving the microscopic constants, which was used by Benesch and Benesch, differed in detail from that outlined here, but was identical in principle.

ratio of these two forms is approximately 1 to 1 rather than 2 to 1. This discrepancy indicates the order of magnitude of the uncertainty involved in the assumption that the effect of an S-alkyl group on the ionization of a neighboring group is the same as that of an SH group.

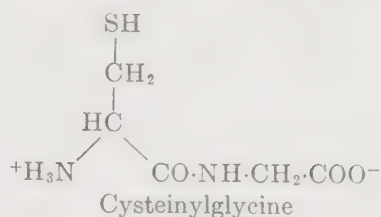
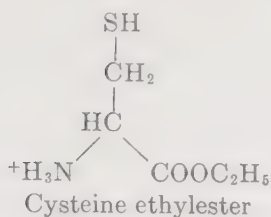
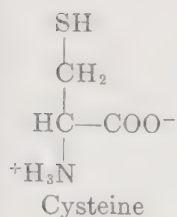
TABLE V
SOME MACROSCOPIC AND MICROSCOPIC pK VALUES OF CERTAIN AMINO-
THIOLS AT 23°

Compound	pK_2	pK_3	pk_{12}	pk_{13}	pk_{123}	pk_{132}	$pH_{(RSH)=(RS^-)}$
L-Cysteine	8.36	10.53	8.53	8.86	10.36	10.03	
L-Cysteine ethyl ester*	6.69	9.17	7.45	6.77	8.41	9.09	
L-Cysteinylglycine	7.07	9.55	7.87	7.14	8.75	9.48	
β -Mercaptoethylamine			8.35	—	—	—	
Glutathione							9.20
L- γ -Glutamylcysteine							9.90
DL-Homocysteine							10.00

The values in the last three rows indicate the pH value at which $\alpha_{SH} = 0.5$ for the substances in question.

* For cysteine ethyl ester, there is no acidic $-\text{COOH}$ group; hence these macroscopic constants should properly be designated as pK_1 and pK_2 , respectively, and the four microscopic constants as pk_2 , pk_3 , pk_{23} , and pk_{32} . From R. E. Benesch and R. Benesch (1955). See text data.

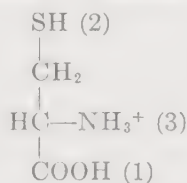
Table V also shows the values of the microscopic constants in the ionization scheme (40.2) for several other sulfhydryl compounds. It will be noted that the pK value of the SH group in L-cysteine ethyl ester is only 7.45, when the neighboring amino group carries a positive charge. This shows how relatively strongly acidic the SH group can become, under the combined influence of the positively charged amino group and the highly polar ester group. In cysteine ethyl ester, however, pk_{13} for the ammonium group is actually lower than pk_{12} for the $-\text{SH}$ group. This inverts the order found in cysteine itself. The reason is readily apparent on considering the structures involved. The carboxyl group of cysteine is negatively charged at neutral or alkaline pH , whereas of course the ethyl ester group is uncharged. This negative charge is closer to the amino group than to the sulfhydryl and therefore has a more pronounced effect in shifting pk_{13} to the alkaline side than in shifting pk_{12} , if we compare the respective values for cysteine and its ethyl ester. Analogous relations for L-cysteinylglycine are apparent from the table. Here, however, a peptide linkage immediately adjoins the charged amino group, and the negatively charged carboxyl has been displaced further away from the other ionizing groups in the peptide, as compared with the amino acid or its ester.



Finally, making use of our previous assumption, based on Wegscheider's work, that a $-\text{COOR}$ group is equivalent to an un-ionized $-\text{COOH}$ group in its effect on the ionization of a neighboring group, we may use the combined data for cysteine and for its ethyl ester, as given in Table V, to calculate a complete microscopic ionization scheme for cysteine. The ionization scheme for cysteine is indeed formally identical with that for glutamic acid already given in Table II. The values of pk_{12} , pk_{13} , pk_{123} , and pk_{132} have already been calculated in the previous discussion. The corresponding four microscopic constants for cysteine ethyl ester from Table V, employing Wegscheider's principle, become equal to k_2 , k_3 , k_{23} , and k_{32} in Table II. We may take $k_1 = K_1$, the first macroscopic dissociation constant, since $K_1 = k_1 + k_2 + k_3$, and we know that k_2 and k_3 are very small compared to k_1 . Here we take $pk_1 = 1.71$ from Chapter 8, Table VI, since we have no value for k_1 at the same ionic strength as for the values in Table V. Then, making use of the interrelations between the constants, as given in Table II, we may derive the complete ionization scheme shown in Table VI. These values are subject to some degree of uncertainty, since they are calculated with the use of assumptions that may not be quite rigorous. They should, however,

TABLE VI
MICROSCOPIC pk VALUES OF CYSTEINE AT 23°

Ionizing group					
Carboxyl		Sulfhydryl		Ammonium	
pk_1	1.71	pk_2	7.45	pk_3	6.77
pk_{21}	2.79	pk_{12}	8.53	pk_{13}	8.86
pk_{31}	3.80	pk_{32}	9.09	pk_{23}	8.41
pk_{231}	4.74	pk_{132}	10.03	pk_{123}	10.36



Values calculated from the data of Table V as explained in the text. The value of 1.71 at 25° is an extrapolated value for zero ionic strength at 25°. It is therefore not quite comparable with the data of Table V and the relations given in Table II to calculate pk_{21} , pk_{31} , and pk_{231} ; all these values are subject to some revision, perhaps of the order of 0.2 in pk .

furnish a useful basis for estimating the types of interaction between ionizing groups to be expected in molecules of related types.

No other polyvalent acids containing three or more acidic groups ($n \geq 3$) have yet been examined in such detail, so far as we are aware. A detailed examination of the ionization of tyrosine could readily be carried out along the same lines as that of cysteine.* For more complex polyvalent acids, the number of microscopic constants becomes very large. Indeed, even for $n = 4$, such a detailed examination would be a rather complex matter. In dealing with polyvalent acids with large values of n it is the general practice to subdivide the acidic groups into a very small number of classes, all the groups in each class being considered inherently identical to simplify the calculations. As the net charge on the acid varies with the gain and loss of protons (or of other ions which may be bound to the polyvalent acid) the electrostatic forces between the groups, and between the polyvalent acid and the surrounding ions, modify the successive dissociation constants. According to the Debye-Hückel theory, these electrostatic effects can be calculated from the dielectric constant of the medium, the ionic strength, and an assumed molecular model for the polyvalent acid—generally taken as a charged sphere (see p. 512). These simplified treatments often yield surprisingly satisfactory results. The complexity of the actual situation, however, should not be forgotten, and the kind of detailed analysis presented here for glutamic acid and cysteine may suggest valuable ways of analyzing certain aspects of the ionization of more complicated molecules.

Isoelectric Points of Amphoteric Substances

Many substances of great importance in biochemistry—notably the amino acids, peptides, and proteins—contain both uncharged and cationic

* Such a study for tyrosine was indeed made by E. Brand, W. H. Goldwater and L. Zucker, *Abstracts 106th Meeting, American Chemical Society*, New York, September 1944, p. 19B; but no full report of their work has appeared. The problem has recently been further investigated by J. T. Edsall, B. R. Hollingworth, and D. B. Wetlaufer, *Abstracts 132 Meeting, American Chemical Society*, New York, September 1957, p. 32C. From their work, the pK' value of the $-\text{NH}_3^+$ group in O-methyl tyrosine, at 25° and ionic strength 0.16, is 9.26; this should be very close to the corresponding microscopic constant for the $-\text{NH}_3^+$ group in tyrosine, when the $-\text{OH}$ group is uncharged and the carboxyl negatively charged. The corresponding value for O-methyl tyrosine ethyl ester is 7.30; that is, the removal of the charge on the carboxyl group shifts the pK' value down by very nearly 2 pH units. This is almost the same shift that is found on comparing the pK_2' of a simple amino acid with that of its ester (see discussion preceding and following equation 25 of this chapter). The full analysis of the microscopic ionization constants of tyrosine involves extensive ultraviolet absorption measurements as a function of pH, between 275 and 300 m μ . The principles are similar to those discussed above for cysteine.

acid groups. The former include such groups as carboxyl, sulfhydryl, phenolic hydroxyl, and partially esterified phosphoric acid residues;⁵ they are uncharged in strongly acid solutions, negatively charged in sufficiently alkaline solutions. The latter include such groups as the conjugate acids of amines and of such organic bases as pyridine, piperidine, imidazole, thiazole, pyrazole, guanidine, and their derivatives generally. In proteins the ammonium, imidazolium, and guanidinium groups are the most important. These are positively charged in strongly acid solution, and the conjugate bases are uncharged in sufficiently alkaline solution. Consider an n -valent acid containing u groups of the uncharged acid type, and c groups of the cationic acid type, so that $n = u + c$. In a sufficiently acid solution all the u groups are uncharged and all the c groups positively charged, so that the net charge, Z , per molecule is $+c$. Here, in the notation previously employed, $h = 0$; that is, the polyvalent acid contains the maximum possible number of bound protons. In a sufficiently alkaline solution all the u groups are negatively charged and all the c groups are uncharged, so that the mean net charge, \bar{Z} , is $-u$. In passing from the former to the latter state $u + c = n$ protons have been removed from the polyvalent acid, so that $h = n$ in strongly alkaline solution. At any stage the mean net charge, \bar{Z} , is equal to $c - \bar{h}$, provided that no ions except protons are released or bound by the ampholyte. In an electric field the molecule migrates as a cation in strongly acid solution, as an anion in strongly alkaline solution. At some intermediate pH value, therefore, the mean net charge, \bar{Z} , must attain the value zero, and the molecule will remain stationary in an electric field. The pH value at which this occurs is known as the isoelectric point.⁶

Experimentally the phenomenon of the isoelectric point was recognized in 1899 by W. B. Hardy, who observed that boiled egg white migrated as a cation in acid and as an anion in alkaline solution. By careful adjustment of the amounts of added acid and alkali he found he could obtain a solution in which the egg white did not move toward either the anode or the cathode. Hardy observed visible particles of denatured egg white, the motion of which was easy to follow by eye; but it was soon found that the same type of behavior was exhibited by undenatured proteins in solution; and the characterization of molecules such as proteins by their isoelectric points has become a matter of great importance.

The mobility of a molecule in an electric field—that is, the velocity

⁵ A monoester of phosphoric acid, $R\cdot OPO(OH)_2$, is a dibasic acid and may be considered as equivalent to two uncharged acid groups.

⁶ If ions other than protons are bound by the ampholyte, then the isoelectric point will of course be affected by such binding (see Chapter 11). Here, however, we restrict ourselves to the cases in which only proton binding need be considered.

of its motion in cm sec^{-1} , divided by the field intensity in volt cm^{-1} —is, in a medium of given ionic strength and dielectric constant, proportional to the mean net charge, \bar{Z} , and inversely proportional to a resistance factor which depends on the size and shape of the molecule and the viscosity of the medium. If all the various forms of a macromolecule have the same size and shape, irrespective of their electric charge, the resistance factor should be the same for all, and then the mobility of each individual form should be proportional to its net charge, if the solvent medium remains essentially unchanged.

Substances which may carry a net charge which is positive, zero, or negative according to the $p\text{H}$ of the solution are often termed amphoteric substances or ampholytes, and this is often a convenient word to use in referring to them. The simplest type of amphoteric substance is a simple amino acid, such as glycine, with $n = 2$ and $u = c = 1$. The calculation of its isoelectric point from its dissociation constants was first given by Michaelis. There is only one cation (A^+) and one anion (A^-) of the amino acid, and at the isoelectric point ($p\text{H}_I$) the concentrations of the two should be equal on the assumptions already stated. This gives

$$(\text{A}^+) = \frac{(\text{H}_I^+)(\text{A})}{K_1} = (\text{A}^-) = \frac{K_2(\text{A})}{(\text{H}_I^+)} \quad (41)$$

where the symbol (A) denotes the total concentration of the two forms with zero net charge—or, for all practical purposes, the concentration of the dipolar ion. From (40) we obtain immediately

$$(\text{H}_I^+)^2 = K_1 K_2 \quad \text{or} \quad p\text{H}_I = \frac{pK_1 + pK_2}{2} \quad (41.1)$$

which gives $p\text{H}_I = (2.3 + 9.7)/2 = 6.0$ for glycine at 25° .

This treatment is strictly valid only at infinite dilution where the activities are equal to the concentrations. At finite concentrations, however, the same equation is valid if we replace K_1 and K_2 by the apparent constants K_1' and K_2' , defined by equations (70) and (71) of Chapter 8. The values of K_1' and K_2' , however, are functions not only of the concentration of the ampholyte itself but of the ionic strength and of the chemical nature of the solvent medium. The K' values used, therefore, must be those valid for the particular state of the system under consideration.

Actually, for most simple ampholytes of this type, such as glycine or glycyglycine, pK_1 and pK_2 are so far apart that there is not merely an isoelectric point, but a broad zone of $p\text{H}$ values in which the ampholyte is practically isoelectric. In glycine, for instance, the concentration of (A^+) and (A^-) is less than 1% of that of (A) at all $p\text{H}$ values between 4.3 and 7.7.

Problem: To calculate the pH of a solution of pure glycine in water at various concentrations—molar, $0.1 M$, $0.01 M$, $10^{-4} M$, $10^{-6} M$, etc.

Set the total glycine concentration equal to C . Hence

$$C = (A^+) + (A) + (A^-)$$

The only ions present are A^+ , A^- , H^+ , and OH^- . For electrical neutrality

$$(H^+) + (A^+) = (A^-) + (OH^-) \quad (42)$$

Rewriting this in terms of K_1 , K_2 , and K_w , we obtain

$$(H^+) \left[1 + \frac{(A)}{K_1} \right] = \frac{K_2(A) + K_w}{(H^+)} \quad (43)$$

or, after rearrangement,

$$(H^+)^2 = \frac{K_1[K_2(A) + K_w]}{(A) + K_1} \quad (44)$$

We know that, as C approaches zero, $pH \rightarrow 7$; and as C becomes large, the pH should approach the isoelectric point (pH 6). Within this pH range, the concentration of the anion (A^-) and the cation (A^+) is less than 0.2% of the concentration of dipolar ion (A), since $K_1 = 10^{-2.3}$ and $K_2 = 10^{-9.7}$. Hence we can set $(A) = C$ in (44) with very little error, and with numerical values inserted for the K 's at 25° , the equation becomes

$$(H^+)^2 = \frac{10^{-2.3}[10^{-9.7}C + 10^{-14}]}{10^{-2.3} + C}$$

If $C \gg 10^{-2.3}$, this reduces to $(H^+) = 10^{-6}$, or $pH = pH_I$; if $C = 0$, $pH = 7$ as it should. Values of (H^+) for intermediate C values are readily calculated. The reader should compute several values and plot pH against $\log C$. Solutions such as the glycine solution just discussed, containing no ions except those of an ampholyte, and hydrogen and hydroxyl ions, are often called isoionic solutions. A further discussion of such solutions is given in Chapter 11.

Isoelectric Points of Polyvalent Ampholytes

The same form of treatment may be readily extended to more complex ampholytes. As the condition for the isoelectric state, we again assume that the average net charge per molecule of ampholyte shall be zero; that is, that the sum of all the positive charges on the ampholyte molecules shall equal the sum of all the negative charges. Here again we are making the assumption that the ratio of mobility to net charge is the same for all the individual ionic species of the ampholyte, in any given medium—in other words, that they all have the same size and shape.

In the general case of an n -valent ampholyte containing u uncharged acid groups and c cationic acid groups, it follows that, when the isoelectric point is attained, the sum of the equivalent concentrations for all the cations must equal the corresponding sum for the anions. The solution contains molecules of class 0A with net charge c , molecules of class 1A with net charge $c - 1$, . . . molecules of class cA with net charge zero, others of class ${}^{c+1}A$ with net charge -1 , . . . up to ${}^{c+u}A$ with net charge $-u$. In general, for a molecular species which has lost h protons, relative to the most acid form 0A , the net charge $Z = c - h$. At the isoelectric point the sum of all the positive charges must be equal to the sum of all the negative charges, and therefore

$$c({}^0A) + (c - 1)(\Sigma {}^1A) + \cdots + 2(\Sigma {}^{c-2}A) + (\Sigma {}^{c-1}A) \\ = (\Sigma {}^{c+1}A) + 2(\Sigma {}^{c+2}A) + \cdots + (u - 1)(\Sigma {}^{c+u-1}A) + u({}^{c+u}A) \quad (45)^7$$

This is the general condition of balance for the isoelectric point, if we assume that no ions other than protons are bound by the ampholyte.⁸ The additional complications that arise when other anions and cations are also bound are discussed in Chapter 11.

In terms of the K 's, as defined by (34) and (35), (45) may be written

$$\frac{c(H_I^+)^c}{K_1 K_2 \cdots K_{c-1} K_c} + \frac{(c - 1)(H_I^+)^{c-1}}{K_2 K_3 \cdots K_c} + \cdots + \frac{2(H_I^+)^2}{K_{c-1} K_c} + \frac{(H_I^+)}{K_c} \\ = \frac{K_{c+1}}{(H_I^+)} + \frac{2K_{c+1}K_{c+2}}{(H_I^+)^2} + \cdots + \frac{uK_{c+1}K_{c+2} \cdots K_{c+u}}{(H_I^+)^u} \quad (46)$$

As a simple application of (46), consider glutamic acid, for which, from Table I, $pK_1 = 2.16$, $pK_2 = 4.32$, $pK_3 = 9.96$. Here $u = 2$, $c = 1$, and equation (46) becomes

$$\frac{(H_I^+)}{K_1} = \frac{K_2}{(H_I^+)} + \frac{2K_2 K_3}{(H_I^+)^2} \quad (46.1)$$

It is clear, however, from Table II, on considering the net charge carried by the different ionic species, that the isoelectric point of glutamic acid (H_I^+) must lie somewhere between K_1 and K_2 . On considering the values of the K 's it is then immediately apparent that the last term on the right in (46.1) is negligible compared to the other two terms. When this is

⁷ The summation signs have been omitted for the species 0A and ${}^{c+u}A$, since, as we have noted before, there is only one microscopic form which corresponds to these particular species.

⁸ We may note that (45) could also be derived from the first equality in (38) by setting $\bar{Z} = c - \bar{h} = 0$ and making use of the relation $n = c + u$, relating the total number of groups to the cationic and uncharged groups.

omitted (46.1) becomes simply

$$(\text{H}_I^+)^2 = K_1 K_2$$

or

$$p\text{H}_I = \frac{(pK_1 + pK_2)}{2} = 3.24$$

We may check this conclusion by inserting $(\text{H}_I^+) = 10^{-3.24}$ as a trial value for (H_I^+) into (46.1), taking the values of the K 's from Table I. It is immediately obvious that the second term on the right is completely negligible compared with the first.

Similarly for lysine, in which $u = 1$ and $c = 2$, the reader may readily convince himself, using equation (46) and the pK_A values given in Table VI of Chapter 8, that the isoelectric point is accurately given by the equation $p\text{H}_I = (pK_2 + pK_3)/2$.

A similar treatment can readily be applied to many ampholytes containing only few ionizing groups. If, as is often true in amino acids and peptides, $K_{c-1} \gg 2(\text{H}_I^+)$, and $(\text{H}_I^+) \gg 2K_{c+2}$ at the isoelectric point, we may neglect all terms of (46) except the last on the left-hand side and the first on the right-hand side. In that case (46) becomes

$$\frac{(\text{H}_I^+)}{K_c} = \frac{K_{c+1}}{(\text{H}_I^+)} \quad \text{or} \quad p\text{H}_I = \frac{pK_c + pK_{c+1}}{2} \quad (47)$$

It is obvious that (46.1) is a special case of (47). Further illustrations of the application of (47) are given in Cohn and Edsall (1943, Chapter 4).

For more complicated ampholytes, it is often convenient to use a very rough approximation procedure to fix the region of $p\text{H}$ in which the isoelectric point is expected to fall. We may illustrate it by an example of a hypothetical peptide; the reasoning can readily be extended to such complex natural products as proteins. Suppose that our peptide contains n_1 free carboxyl groups, all alike and ionizing independently, with a titration constant, pG_1 , of (say) 4.3, n_2 imidazole groups of histidine residues, each with a pG_2 of 7.0, and n_3 ϵ -ammonium groups of lysine residues, each with pG_3 of 10.0. The number of negatively charged $-\text{COO}^-$ groups at any $p\text{H}$ is then $n_1\alpha_1$, where α_1 is calculated from (27), with $pG_a = 4.3$. The number of positively charged imidazolium groups is $n_2(1 - \alpha_2)$, where α_2 is also calculated from (27), with $pG_a = pG_2 = 7.0$; and similarly the number of positively charged ammonium groups is $n_3(1 - \alpha_3)$. Therefore the total mean net charge is

$$\begin{aligned} \bar{Z} &= n_2(1 - \alpha_2) + n_3(1 - \alpha_3) - n_1\alpha_1 \\ &= n_2 + n_3 - (n_1\alpha_1 + n_2\alpha_2 + n_3\alpha_3) = c - \Sigma n_i\alpha_i = c - \bar{h} \end{aligned} \quad (48)$$

For example, suppose $n_1 = 20$ carboxyl groups, $n_2 = 10$ imidazole groups, and $n_3 = 15$ amino groups; hence $c = 25$. Then it is plain that at the isoelectric point practically all the carboxyls will be negatively charged, and practically all the amino groups positively charged. Therefore, on the average, 5 of the 10 imidazole groups must carry positive charges to make the net charge zero. This means that $\alpha_1 \cong 1$, $\alpha_3 \cong 0$, and $\alpha_2 = 0.5$; hence $pH_I = pG_2 = 7$. If we took $n_1 = 30$, $n_2 = 10$, $n_3 = 15$, and $c = 25$, the same considerations would give $\alpha_2 \cong \alpha_3 \cong 0$; $\alpha_1 = 0.833$, and $pH_I = 4.3 + \log [\alpha_1 / (1 - \alpha_1)] = 5.0$. On the other hand, if $n_1 = 15$, $n_2 = 10$, $n_3 = 30$, and $c = 40$, the condition for $\bar{Z} = 0$ would give $\alpha_1 \cong \alpha_2 \cong 1$; $\alpha_3 = 0.5$, and pH_I would be 10. Naturally this kind of calculation is a rough one, for actual titration curves of complex polybasic acids and ampholytes cannot be fitted by the simple equations (27), (28), and (29); deviations are caused by the electrostatic interactions discussed below, and also by the fact that not all the groups of a given type have the same intrinsic acid strength. Given an accurate amino acid analysis for a protein, however, a calculation along the lines of equation (48) usually suffices to give a reasonable approximation to the expected isoelectric point. If the experimentally observed isoelectric point deviates markedly from that so calculated, three possible explanations may be considered: (1) the amino acid analysis is inaccurate; (2) the protein is selectively binding anions or cations, other than the hydrogen ion, which are present in the solution; or (3) some of the groups may be involved in intramolecular hydrogen bonds, or other attachments, so that they become nontitratable, or so that their pK values are greatly displaced from the usual values characteristic of such groups. The binding of ions other than H^+ ions by macromolecules is considered in Chapter 11. The effects of internal hydrogen bonding are discussed briefly later in this chapter; the reader may also be referred to an interesting paper by Laskowski and Scheraga (1954).

Equilibrium between Different Ionic Forms in Polyvalent Acids and Ampholytes: The Equation of Linderström-Lang

It is apparent that in a polyvalent acid or ampholyte the mean net charge, \bar{Z} , is truly an average over a large number of individual forms which are all in a mobile equilibrium with one another. Any individual molecule is constantly giving up and taking on protons, so that its actual net charge, Z , at any moment may be greater or less than \bar{Z} . If the molecules are all alike, however, the value of Z , averaged over a considerable period of time for any one molecule, will be equal to \bar{Z} obtained by averaging over the whole system of molecules in the solution at any instant. At

the isoelectric point of an ampholyte ($\bar{Z} = 0$) there must be anions and cations of the ampholyte in the solution at any moment, together with the forms of zero net charge. In a simple amino acid like glycine, for which $K_1 \gg (H^+)$ and $(H^+) \gg K_2$, the number of anions and cations is very small compared to the number of dipolar ions, and for many purposes may be neglected. If, however, some of the pK values of the ampholyte lie close to pH_I , the number of ions carrying a net charge is relatively large. This is important, for it means that, in terms of long-range electrostatic interactions with neighboring ions and molecules, a multivalent ampholyte such as a protein, even at its isoelectric point, may behave more like an ion carrying a net charge than like a simple dipolar ion. (See the discussion in Chapter 5.) It is also of importance for the dielectric properties of such molecules (Chapter 6), especially in considering their motion in alternating electric fields. The simplest index of the relative importance of these charged forms is the mean square value of the net charge, \bar{Z}^2 . The difference, $\bar{Z}^2 - (\bar{Z})^2$, is the standard deviation of the net charge, which gives a good measure of the spread of the net charge distribution. It has been shown by Linderstrøm-Lang (see Cohn and Edsall, 1943, p. 462) that this quantity is directly proportional to the slope of the titration curve at any point. This result is obtained by differentiating (38) with respect to pH (see also the discussion preceding equation 45):

$$\begin{aligned} \frac{\partial \bar{h}}{\partial pH} &= - \frac{\partial \bar{Z}}{\partial pH} = -2.303 \frac{\partial \bar{h}}{\partial \ln (H^+)} = -2.303 (H^+) \frac{\partial \bar{h}}{\partial (H^+)} \\ &= 2.303 \left[\sum \left(\frac{L_h}{(H^+)^h} \right) \cdot \sum \frac{h^2 L_h}{(H^+)^h} - \left(\sum \frac{h L_h}{(H^+)^h} \right)^2 \right] / \left(\sum \frac{L_h}{(H^+)^h} \right)^2 \\ &= 2.303 [\bar{h}^2 - (\bar{h})^2] = -2.303 [\bar{Z}^2 - (\bar{Z})^2] \quad (49) \end{aligned}$$

We may note that $\partial \bar{h} / \partial pH$ is equal to the "buffer value" of Van Slyke (Chapter 8, equation 37) expressed per mole of polyvalent acid.

As one example we may take the case of horse hemoglobin. (Human hemoglobin is very similar.) From the titration reported by Cohn, Green, and Blanchard (1937) the net charge of the hemoglobin molecule near its isoelectric point changes by approximately 8 proton units per pH unit. From (49) we then obtain $\bar{Z}^2 - (\bar{Z})^2 \approx 3.5$; and at the isoelectric point, where $\bar{Z} = 0$, this becomes simply $\bar{Z}^2 \approx 3.5$. This gives the root mean square value of the net charge per molecule as 1.87. A more detailed calculation of the distribution of charge (Cohn and Edsall, 1943, pp. 460-468) shows that only about 22% of all the hemoglobin molecules are actually isoelectric *at any moment*, when the solution is at its isoelectric point.

Electrostatic Effects on Ionization in Polybasic Acids. The Charged Sphere Model

Actual polybasic acids do not conform to the simple relation, embodied in equation (40), for a set of equivalent *and independent* ionizing groups. There are many cases, however, in which a set of groups, although they may be regarded as essentially equivalent, cannot be considered independent, since the removal of a proton from one group, by making the net charge on the molecule more negative, increases the work required to remove a proton from one of the other groups. Even in such complicated molecules as proteins, it is often possible to describe the titration curve in terms of a few classes of groups, each group being characterized by a single titration pG value—that is, a constant which is taken as characteristic of that group in the absence of electrostatic forces, and is identical with pG in (27). The actual course of the titration is then analyzed in terms of the number of acidic groups in each class, the intrinsic pG value for each class, and the effect of the electrostatic forces in modifying the “ideal” titration curve. The latter, which is described by equations (27) and (29) in terms of the number of groups in each class, the titration pG value for each class, and the pH , should correspond to the titration curve in a hypothetical medium of infinite dielectric constant, in which the electrostatic forces vanish. Equation (40) gives the relations between the successive dissociation constants (K values) corresponding to this single pG value in each class of group.

The simplest model of a macromolecule to assume for electrostatic calculations is a uniformly charged sphere. For many of the “globular” proteins, a spherical model is a reasonably plausible first approximation to the actual structure, although it would be a very bad model for fibrous proteins or nucleic acids, or for the synthetic polyelectrolytes which will later be mentioned briefly. If the molecular weight of a spherical macromolecule (including bound solvent) is M , and if its specific volume is \bar{v} , then the volume of a single molecule is $M\bar{v}/N$. If the molecule is considered as a sphere of radius b , the volume is $4\pi b^3/3 = M\bar{v}/N$, from which the radius may be readily calculated. Thus serum albumin, with $M = 66,000$ approximately, and a partial specific volume for the anhydrous protein of 0.733, has been assumed in many calculations to bind about 0.2 gram of water per gram of protein, although this figure is somewhat arbitrary and uncertain. With this figure, and with the bound water assumed to have a specific volume of 1.00, the radius b of the sphere taken as representing the serum albumin molecule is very nearly

$$30 \times 10^{-8} \text{ cm} = 30 \text{ \AA}$$

It may appear arbitrary to treat the sphere as if its charge were spread uniformly over its surface, for actually the charge is made up of a number of discrete charged groups, each having a more or less fixed position on the surface of the sphere. The total concentration of molecules of any given net charge, however, is made up of many different microscopic forms, as we have already shown in detail. The average distribution for all these microscopic forms should be much closer to a uniform spreading of charge over the surface of the molecule than would be the charge distribution for any one of these forms. The main justification, however, for the assumption of a uniformly charged sphere as a model is that the

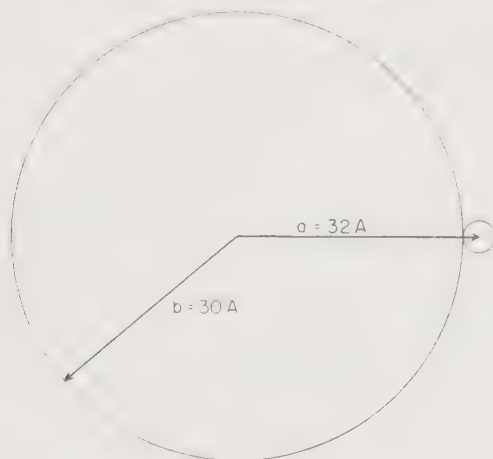


FIG. 7. A macro-ion of radius 30 Å, and a small ion of radius 2 Å, showing the values of a and b in equation (53) for this particular case.

model works reasonably well for many systems—works, indeed, much better than we should have any right to expect.

Let Z be the net charge, in proton units, on the sphere of radius b ; hence $Z\epsilon$ is its charge in electrostatic units. Let a be the “collision diameter, which represents the closest distance of approach of the center of a neighboring small ion to the center of the sphere (Fig. 7). For small ions such as sodium and chloride, it is reasonable to take a about 2 Å greater than b , or 32 Å for serum albumin. If the sphere is immersed in a medium of dielectric constant D , at an ionic strength ω , corresponding to a reciprocal distance of the ionic atmosphere $\kappa = 3.286 \times 10^7 \sqrt{\omega} \text{ cm}^{-1}$, then from Chapter 5, equation (72.1), the electrical free energy of the sphere is

$$\bar{F}_e = \frac{Z^2 \epsilon^2}{2D} \left(\frac{1}{b} - \frac{\kappa}{1 + \kappa a} \right) \quad (50)$$

We note that $Z = c - h$, where c is the maximum number of cationic acid groups (see the discussion on isoelectric points), and h is the total

number of acid groups removed from the most acid form of the polyvalent acid.

The logarithm of the activity coefficient (γ_h) for an ion of class h is conveniently defined so that $\gamma_h = 1$ and $\log \gamma_h = 0$ in the absence of electrostatic forces—that is, at zero net charge or in a hypothetical medium of infinite dielectric constant. Then, for an ion of net charge $Z = c - h$,

$$\ln \gamma_h = \frac{(\bar{F}_e)_h}{kT} = \frac{(c - h)^2 \epsilon^2}{2DkT} \left(\frac{1}{b} - \frac{\kappa}{1 + \kappa a} \right) = (c - h)^2 w = Z^2 w \quad (51)$$

or

$$\gamma_h = e^{(c-h)^2 w} = e^{Z^2 w} \quad (52)$$

Here the parameter w is defined by the equation

$$w = \frac{\epsilon^2}{2DkT} \left(\frac{1}{b} - \frac{\kappa}{1 + \kappa a} \right) \quad (53)$$

If the radius b is 30×10^{-8} cm, then w at zero ionic strength in water at 25° is 0.119. If the ionic strength is 0.1 ($\kappa = 1.037 \times 10^7$ cm $^{-1}$) and $a = 32 \times 10^{-8}$ cm, then w is 0.0334. The value of w decreases progressively as the ionic strength increases, but, since $a > b$, w always remains positive, as may be seen by setting $\kappa = \infty$ in (53).

Consider now a set of identical acidic groups in the spherical macromolecule, each of intrinsic constant k_{int} . Here k_{int} is equivalent to the titration constant, G , of equations (40), except that we are now considering the effects of electrostatic interactions, which were ignored in (40). Suppose, for instance, that the groups are carboxyls, and furthermore that they are the most strongly acidic groups in the molecule, which are therefore the first to be titrated if we begin the titration in strongly acid solution and add OH^- ion. We further assume for simplicity that all the carboxyl groups lose their protons before any of the other acidic groups in the molecule start to lose theirs; thus the molecule contains a fixed positive charge (c) due to the cationic groups present, and a variable negative charge, h , equal to the number of carboxyl groups which have lost protons. The net charge is $Z = c - h$. At the start of the titration, Z is equal to c , the number of cationic acid groups present in the molecule, and $h = 0$. The first dissociation constant, K_1 , is from (40) equal to $nG = nk_{\text{int}}$. From equation (31) but with activity coefficients inserted, it may be written

$$\begin{aligned} K_1 &= \frac{(\text{H}^+)(\text{activity of ions of class } {}^1A)}{(\text{activity of ions of class } {}^0A)} = \frac{(\text{H}^+)(\Sigma {}^1A)\gamma_1}{({}^0A)\gamma_0} \\ &= nk_{\text{int}} = \frac{(\text{H}^+)(\Sigma {}^1A)e^{(c-1)^2 w}}{({}^0A)e^{c^2 w}} = \frac{(\text{H}^+)(\Sigma {}^1A)}{({}^0A)} e^{(-2c+1)w} \quad (54) \end{aligned}$$

In equation (54) the values of the activity coefficients γ_1 and γ_0 have been inserted from (52). Thus the total concentration of all species of the class 1A is given by

$$(\Sigma ^1A) = \frac{K_1}{(H^+)} \frac{\gamma_0}{\gamma_1} (^0A) = \frac{nk_{\text{int}}}{(H^+)} e^{(2c-1)w} (^0A) \quad (54.1)$$

and in general, by making use of (40) and (52), the *concentration* of all species of the class hA is given by

$$(\Sigma ^hA) = \frac{K_1 K_2 \cdots K_h}{(H^+)^h} \frac{\gamma_0}{\gamma_h} (^0A) = \frac{n!}{(n-h)!h!} \frac{k_{\text{int}}^h}{(H^+)^h} e^{(2ch-h^2)w} (^0A) \quad (54.2)$$

Thus the titration curve, based on these activity coefficients, is given by a modified form of (38) and (40):

$$\bar{h} = \frac{\sum_{h=0}^n \frac{hL_h\gamma_0}{(H^+)^h\gamma_h}}{\sum_{h=0}^n \frac{L_h}{(H^+)^h\gamma_h}} = \frac{\sum_{h=0}^n \frac{n!}{(n-h)!h!} \frac{hk_{\text{int}}^h}{(H^+)^h} e^{(2ch-h^2)w}}{\sum_{h=0}^n \frac{n!}{(n-h)!h!} \frac{k_{\text{int}}^h}{(H^+)^h} e^{(2ch-h^2)w}}} \quad (55)$$

The exact computation of the titration on curves defined by (55) is a very tedious job, even if the total number of titratable groups, n , is known in advance. Qualitatively the curve is not unlike a simple curve for n independent groups. It is, however, spread more widely on the pH axis; the larger the value of w , the less steep the curve is at the mid-point, where $\bar{h} = n/2$. The slope at the mid-point was calculated by Linderstrøm-Lang (1924), in a pioneer investigation, and was found to be

$$\left(\frac{d\bar{h}}{dpH} \right)_{\text{mid}} = - \left(\frac{d\bar{Z}}{dpH} \right)_{\text{mid}} = \frac{2.303n}{2(wn+2)} \quad (56)$$

The slope of the "ideal" curve for $w = 0$ at the mid-point is therefore equal to $2.303n/4$, in agreement with the simple relation to be expected (Chapter 8, equation 37) when $\alpha = 1 - \alpha = 0.5$. At the mid-point of the curve, when half the n titratable groups have lost their protons, $\alpha = \bar{h}/n = 0.5$. Near the mid-point the titration curve is very nearly a straight line, and we may write from (56), comparing the slope of the actual curve (at finite w) with the "ideal" curve ($w = 0$):

$$\left(\frac{dpH}{d\bar{Z}} \right)_w - \left(\frac{dpH}{d\bar{Z}} \right)_{w=0} = 0.868 \left(\frac{wn+2}{n} - \frac{2}{n} \right) = -0.868w \quad (57)$$

This relation in the neighborhood of the mid-point suggested a simple form for the approximate solution of (55), which has proved to fit the exact solution very closely indeed, as was first pointed out by Cannan *et al.* (1942). The "ideal" curve for $w = 0$ is of course identical with that of n equivalents of a univalent acid with titration constant $pG = pk_{\text{int}}$, as given by (40). When a set of curves corresponding with various positive values of w is compared with such an ideal curve, it is found empirically that the points on the former curve, at any fixed value of \bar{h} , are displaced from the latter curve along the pH axis by

$$0.868w\bar{Z} = 0.868w(c - \bar{h})$$

This relation, for the central linear portion of the curve near $\bar{h} = n/2$, follows from (57). In fact the relation holds well over the whole course of a series of curves corresponding to a wide range of values of w —a wider range than has been employed in the actual study of proteins and other polyvalent ampholytes. This relation leads to the following approximate equations for a set of n acidic groups, all of which are assumed to have the same intrinsic constant pk_{int} :

$$pH = pk_{\text{int}} + \log [\alpha/(1 - \alpha)] - 0.868w\bar{Z} \quad (58)$$

$$pH = pk_{\text{int}} + \log [\alpha/(1 - \alpha)] - 0.868w(c - \alpha n) \quad (59)$$

$$pH = pk_{\text{int}} + \log [\bar{h}/(n - \bar{h})] - 0.868wc + 0.868w\bar{h} \quad (60)$$

Here α represents the fraction of all the groups of the set which have lost protons; that is, $\alpha = \bar{h}/n$. It is obvious from (58) that the ideal curve for n equivalent and independent groups is immediately obtained if $w = 0$. The deviation from the ideal curve, due to the effect of the term involving w , can be resolved into two parts. The presence of the c cationic charges displaces the entire curve to the left along the pH axis, by an amount equal to $-0.868wc$ pH units, but does not change the *form* of the curve. The effect of the variable charge on the n groups under consideration is given by the term $+0.868w\bar{h} = +0.868w\alpha n$. This spreads the curve out more widely over the pH axis than the ideal curve would be spread, so that the buffer value is less for the actual than for the ideal curve (see equation 56).

Equation (58) is an enormous simplification as compared with (55), and the curves given by (58) fit those derived from (55) within 0.1 equivalent in \bar{h} , if n is more than 4 or 5. For polyvalent acids with many groups, this discrepancy is trivial compared with the experimental error. The effect of the electrostatic factor, w , and of the presence of a fixed cationic charge, c , is illustrated in Fig. 8. The number of carboxyl groups, n , is taken as 58, and the value of pk_{int} as 4.6; these correspond to the values

assumed by Cannan *et al.* (1942) for β -lactoglobulin. The dotted curve (c) represents the ideal curve for titration of 58 groups of $pK_{int} = 4.6$ in the absence of electrostatic forces or of any fixed cationic charge. Curve b shows the effect on the titration of electrostatic forces, if the factor w is taken as 0.03. If 45 positively charged groups are then added to the molecule ($c = 45$), at the same value of w , curve a is obtained. This curve is identical in form with curve c, but is displaced to the left (see equation 58) by $-0.868wc = -1.15$ pH units. Curve d illustrates the effect of increasing w from 0.03 to 0.10—that is, of *decreasing* the ionic strength from a value near 0.1 to a value near 0.01 (taking b as near

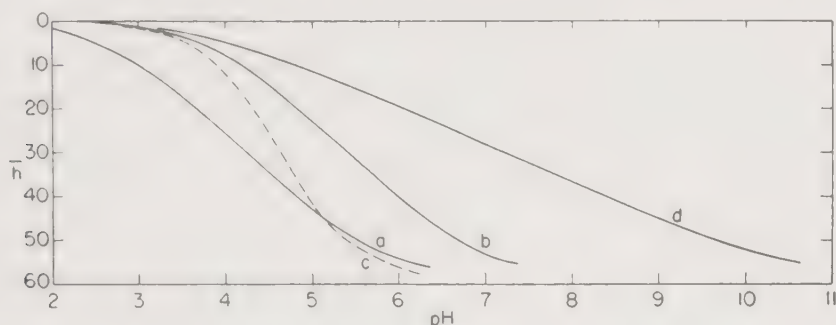


FIG. 8. Titration curves for various values of the electrostatic factor w , for a molecule containing 58 equivalent carboxyl groups, each with a pK value of 4.6. This corresponds to the value assumed for the carboxyl groups of β -lactoglobulin by Cannan, Palmer, and Kibrick (1942). Curve a is for $w = 0.03$, when the molecule contains a fixed positive charge $c = 45$ proton units, due to cationic acid groups. Curve b is for $w = 0.03$, $c = 0$. Note that curves a and b are parallel, the fixed positive charge causing a displacement of the entire curve to the left. Curve c (dotted line) is the ideal curve ($w = 0$) for 58 equivalent groups in the absence of interactions. Curve d shows the effect of increasing w to 0.10, taking $c = 0$. Note the greatly increased spread of this curve as compared with curve b.

30×10^{-8} cm). Curve d, like curve b, is drawn on the assumption that no fixed cationic charge is present ($c = 0$). Comparison of the two curves shows how the increase of the electrostatic factor spreads the titration out over a much wider range of pH values.

The situation is essentially the same if we consider the titration of the c cationic groups, in a more alkaline pH range. At high pH values all the n carboxyl groups have been converted into their conjugate bases, and are negatively charged. Hence the titration of the c cationic groups, which again we assume to be all equivalent, is carried out in the presence of a fixed *negative* charge, $-n$. The fraction of the cationic groups from which protons have been removed is α , and the number of positive charges remaining is $(1 - \alpha)c$. Thus in this case $\bar{h} = n + \alpha c$, and

$$\bar{Z} = c - \bar{h} = -n + (1 - \alpha)c$$

Equation (58) still applies, if we take the characteristic value of pk_{int} as the value for the c cationic groups—if the groups are ϵ -ammonium groups of lysine at 25°; for example, pk_{int} is generally in the range 10.1 to 10.6. Equations (59) and (60) become, respectively,

$$pH = pk_{\text{int}} + \log [\alpha/(1 - \alpha)] + 0.868wn - 0.868w(1 - \alpha)c \quad (59.1)$$

$$pH = pk_{\text{int}} + \log [(\bar{h} - n)/(c - \bar{h} + n)] + 0.868wn - 0.868w(1 - \alpha)c \quad (60.1)$$

In actual molecules, such as proteins, there are regions in which the titration of different kinds of groups overlaps. For instance, in the pH region 5.5 to 6, positively charged imidazole groups begin to lose protons before the carboxyl groups are completely ionized. In the region around pH 10 there is often extensive overlapping between the hydroxyl groups of tyrosine and the ϵ -ammonium groups of lysine residues. It is reasonable, however, in such cases to try the application of equation (58) to the titration of any set of similar groups in such a complex molecule, making a preliminary estimate of \bar{Z} from the titration curve itself, taking $\bar{Z} = 0$ at the isoelectric point. This preliminary value of \bar{Z} may then be refined by taking account of the binding of other ions present in the system, using the methods which are described in Chapter 11.

Formulation in Terms of Association Constants

It is often convenient to formulate equilibria in terms of association rather than dissociation constants. For acid-base equilibria this means that we begin by considering the most basic form of the substance under study and then consider the number of protons added, rather than starting with the most acid form and considering the number of protons removed. For a monovalent base B , which can bind a proton to give the conjugate acid HB , the association constant, k_{assoc} is simply the reciprocal of the acid constant, K_A :

$$k_{\text{assoc}} = \frac{(HB)}{(B)(H^+)} = \frac{1}{K_A} \quad (61)$$

The term k_{assoc} has been called by Brönsted the “basicity constant” of the base B ; it should not be confused with the “basic dissociation constant,” K_B , defined in Chapter 8, equations (5) and (7), which is a quite different quantity.

Similarly for a polyvalent ampholyte with a set of n equivalent, but in general not independent, groups we can define an intrinsic association constant for protons, which we denote by $k_{\text{assoc}} = 1/k_{\text{int}}$; hence

$$\log k_{\text{assoc}} = pk_{\text{int}}$$

We can also define the state of ionization of the acid or ampholyte by taking its most basic form as our point of reference, and considering the average number of protons bound per molecule of polyvalent acid—a quantity which we denote by $\bar{\nu}_H$. If we consider the same set of n equivalent groups which we have discussed above, again in the presence of the fixed cation charge, c , it is apparent that $\bar{\nu}_H = 0$ when all n protons of the class under consideration have been removed—that is, in general $\bar{\nu}_H = n - \bar{h}$. Making use of these relations we can rewrite (58), which becomes

$$pH = \log k_{\text{assoc}} - \log \frac{\bar{\nu}_H}{n - \bar{\nu}_H} - 0.868w\bar{Z} \quad (62)$$

If we multiply (62) by 2.303 to convert to natural logarithms, and then take the exponentials of the quantities involved, we obtain, after rearrangement,

$$\frac{\bar{\nu}_H}{n - \bar{\nu}_H} = k_{\text{assoc}}(H^+)e^{-2w\bar{Z}} \quad (63)$$

When solved for $\bar{\nu}_H$ this becomes

$$\bar{\nu}_H = \frac{nk_{\text{assoc}}(H^+)e^{-2w\bar{Z}}}{1 + k_{\text{assoc}}(H^+)e^{-2w\bar{Z}}} \quad (64)$$

Equation (64) is applicable not only for the binding of hydrogen ions but for the binding of any univalent cation to a spherical macromolecule of mean net charge, \bar{Z} , if we are justified in assuming that the charge acts as if it were uniformly distributed over the surface of the sphere, and if there are n equivalent sites at which the cation may be bound. More generally, if we are dealing with the binding of a small ion (A) of charge Z_A to a spherical macromolecule of mean net charge \bar{Z}_p , we may write

$$\bar{\nu}_i = \frac{nk_{\text{assoc}}(A)e^{-2w\bar{Z}_pZ_A}}{1 + k_{\text{assoc}}(A)e^{-2w\bar{Z}_pZ_A}} \quad (65)$$

It is clear from (65) that the electrostatic term favors association of the ion A if \bar{Z}_p and Z_A are of opposite sign, and hinders association if they have the same sign. Qualitatively, of course, this is intuitively obvious. Equation (64) is a special case of (65).

The justification for equation (65) may be seen by means of an argument similar to that given by Scatchard (1949) and by Scatchard *et al.* (1950). Consider a spherical protein molecule, P, which interacts with an ion, A, of valence Z_A . We assume as before that P contains n equivalent sites, each of which can combine with A. Consider a particular configura-

tion, which we denote by $(PA_\nu)_i$, in which the $\nu(A)$ ions are combined at a set of *specified* sites with P. Let the valence of $(PA_\nu)_i$ be Z_p . Consider the equilibrium involved when one ion of A combines with $(PA_\nu)_i$ at one particular vacant position, denoted by j , to form $(PA_{\nu+1})_j$ with net charge $Z_p + Z_A$. From equation (52) the activity coefficient of this latter is $e^{(Z_p+Z_A)^2w}$, and that of the former is $e^{Z_p^2w}$. Thus the equilibrium constant for the reaction



is given by

$$k_{\text{assoc}} = \frac{(PA_{\nu+1})_j e^{(Z_A+Z_p)^2w}}{(PA_\nu)_i(A) e^{Z_p^2w}} = \frac{(PA_{\nu+1})_j e^{(2Z_pZ_A+1)w}}{(PA_\nu)_i(A)} \cong \frac{(PA_{\nu+1})_j e^{2Z_pZ_Aw}}{(PA_\nu)_i(A)} \quad (66)$$

We recall that w is small, generally much less than 0.1. The approximation of setting $2Z_pZ_A + 1$ equal to $2Z_pZ_A$ in the exponential is obviously reasonable if Z_p is large. If Z_p is near zero, the coefficient $2Z_pZ_Aw + 1$ is so small that its exponential is very close to unity, and can practically be neglected in (66).

The fraction of the molecules for which the site j is occupied by A—or, which is the same thing, the probability that site j is occupied by A in any molecule, of the configuration specified above, chosen at random from the solution—is given by

$$p_j = \frac{(PA_{\nu+1})_j}{(PA_\nu)_i + (PA_{\nu+1})_j} = \frac{k_{\text{assoc}}(A) e^{-2Z_pZ_Aw}}{1 + k_{\text{assoc}}(A) e^{-2Z_pZ_Aw}} \quad (67)$$

If we have n equivalent sites, then the mean number of sites occupied is $\bar{\nu} = np_j$; and the *mean* net charge on the protein is \bar{Z}_p . If we substitute \bar{Z}_p for Z_p in (67) and multiply both sides by n , (65) is obtained. This argument is not to be taken as a rigorous demonstration but as an indication of the line of reasoning involved.

In an actual solution, a protein, nucleic acid, or other macromolecule is surrounded in general by a variety of ions. The net charge, \bar{Z}_p , is determined by the binding of *all* these ions, so that we may write

$$\bar{Z}_p = \sum_{\text{all ions}} \bar{\nu}_i Z_i \quad (68)$$

It is often convenient to take the isoelectric point of the protein in the absence of salts as the reference point ($\bar{Z}_p = 0$) in (68). In that case we assume that $\bar{\nu}_H$ becomes negative when hydroxyl ion is added to the isoelectric protein, positive if hydrogen ion is added, but the assumption that salt is absent in the reference state means that $\bar{\nu}_i$ can assume only zero or positive values for all other ions.

The binding of ions other than protons by macromolecules is discussed in more detail in Chapter 11. A few remarks on such binding may, however, be offered here. Very few proteins show significant binding of the cations of the alkali metals, although myosin does bind both sodium and potassium ions to a significant degree. Most proteins, near their isoelectric points, also show little or no binding of the simple monovalent anions, such as halide ions. Some proteins, however, notably the serum albumins, show a remarkable tendency to bind almost all kinds of anions; as might be inferred from (65), the binding of anions increases as the total net charge on the protein becomes more positive, that is in acid solutions. Serum albumin, however, is found to bind significant amounts of many anions, even when its total net charge is negative. Serum albumin shows strong binding of many anionic organic dyes, and extremely strong binding of long-chain fatty acid anions, and many other proteins show a strong tendency to bind such anions also. Divalent cations, such as the alkaline earths, conversely are bound most strongly to proteins carrying a net negative charge and are often bound weakly or not at all to proteins at or acid to their isoelectric points. Cations of the transition elements, such as Mn^{++} , Cu^{++} , and Zn^{++} , are often bound much more strongly than the alkaline earths, and their interactions give rise to special problems, which are further discussed in Chapter 11. Calculation of \bar{Z}_p in a solution containing several kinds of ions which can be bound to the macromolecule under study requires careful and sometimes tedious computations, since the binding of each ion is from (65) a function of the total net charge on the macromolecule, and the net charge is from (68) a function of all the ions present which are bound. In most cases it is necessary to proceed by successive approximations in calculating \bar{Z}_p . Generally the results are easiest to interpret if only such ions as sodium (or potassium) and chloride are present in the solution, in addition to hydrogen and hydroxyl ions, and the ions of the macromolecule itself.

We have here developed the picture of a macromolecular ion as a sphere with a specified net charge uniformly distributed over its surface. Moreover, we have tacitly supposed that the charge remains uniformly distributed over the surface even when another charged body approaches the sphere. In other words the charged sphere is assumed to be non-polarizable. The effective dielectric constant for electrostatic interactions (see the discussion in Chapter 8, p. 461) is taken to be identical with the measured dielectric constant of the pure solvent in the absence of salt. All these assumptions are open to serious question. Protein molecules, even those commonly referred to as "globular," are probably not generally spherical. When they carry a net charge, this charge is not spread uniformly over the surface but is concentrated at a number of specific

points where the charged groups are located. The assumption that the sphere carries a charge uniformly distributed over its surface appears almost incompatible with the assumption that the sphere is nonpolarizable. In view of all these criticisms it may seem extremely surprising that the spherical model proposed here should work well at all. Nevertheless it is a fact that this model has proved extremely useful, as we shall see below. In view of the limitations and the artificiality of the model, however, it is often best to look upon the electrostatic factor, w , in equations (53) and (58) as an empirical parameter to be determined by the experimental data, rather than as the radius of a spherical molecule. Equation (58) does, in fact, give a very good fit to the titration data for a number of proteins, at least over a limited range of pH and ionic strength, and when this is the case w may be calculated from the slope of the titration curve. Actually in many cases the values of w so obtained give very plausible values for the radius of a spherical molecule of the same molecular weight and partial specific volume as the protein under study. It would not be expected, of course, that very elongated molecules, such as myosin or the nucleic acids, could be represented properly by a spherical model. Such molecules could better be represented, for instance, by cylinders; and electrostatic calculations for a cylindrical model have been given recently, for instance by Hill (1955). We proceed now, however, to discuss the experimental data on certain globular proteins and to their interpretation in terms of the model which is described. We shall begin by discussing the titration curve of β -lactoglobulin, in which complexities are at a minimum.

Acid-Base Equilibria in β -Lactoglobulin Solutions

β -Lactoglobulin can be separated from milk as large and well-formed crystals. An excellent summary of its preparation and properties has been given by McMeekin (1954). Its molecular weight is near 35,000 to 37,000, and its isoelectric point is at 5.18. Its amino acid composition is given in Chapter 3, Table II. It is a typical globulin, only very slightly soluble in pure water at the isoelectric point, but increasing rapidly in solubility as moderate amounts of salt are added (Chapter 5, Fig. 21).

A very careful study of the titration curve of this protein has been made by Cannan *et al.* (1942). They determined titration curves at eight different ionic strengths, ranging from 0.01 to 2.1. Three of these curves are shown in Fig. 9. The pH of the isoelectric point was found to be independent of ionic strength and equal to 5.18 in all the solutions studied. It is apparent from the figure that the curves become steeper, near the isoelectric point, as the ionic strength is increased. This is exactly what would be predicted from (58), since increase of ionic strength

decreases w , as is apparent from (53), and thereby reduces the electrostatic interactions which tend to flatten the curve.

Much can be learned from the titration curve concerning the nature of the ionizing groups in the molecule. Cannan *et al.* assumed a molecular weight of 40,000, and we shall follow them in using this figure, although it is slightly higher than the best values available today, which are near 36,000. At pH values in the range 1.5 to 2, the free carboxyl groups should all exist as uncharged $COOH$ groups, and all cationic groups should be positively charged. Therefore the number of protons bound, on proceeding from the isoelectric point ($\bar{Z} = 0$) to saturation with acid near pH 2,

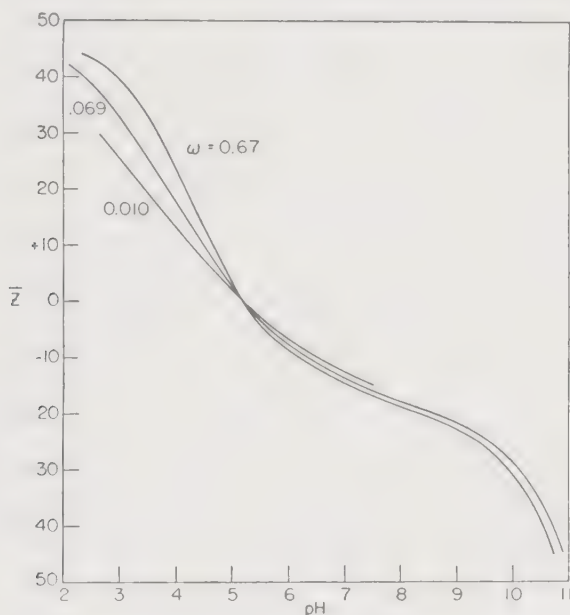


FIG. 9. Titration curves of β -lactoglobulin at three different ionic strengths. (From the data of Cannan, Palmer, and Kibrick, 1942.)

should be equal to the total number of cationic groups in the molecule—that is, to the sum of the α -amino, ϵ -amino, imidazole, and guanidinium groups. This proved to be equal to 46, the maximum value of \bar{Z} .

There is a transition zone in the curves—that is, a region of relatively weak buffering power between two steeper portions of the curves—near pH 8.5. If we take into account the characteristic pK values of the various groups involved (see, for instance, the data of Chapter 8, Table VI), it may be expected that this should correspond to the region in which all the carboxyl groups have acquired negative charges, and the imidazole and α -amino groups have lost their positive charges, while the ϵ -ammonium groups of the lysyl residues and the guanidinium groups of the arginyl residues are still positively charged. Thus the net charge near

pH 8.5 should correspond to $(Lys) + (Arg) - (COOH)$, where $(COOH)$ denotes the sum of all the ionizable carboxyl groups. The observed value of \bar{Z} at this point is -18 to -19 .

The number of lysyl amino groups was estimated by observing the effect of added formaldehyde on the titration curve. Formaldehyde reacts with uncharged amino groups to form hydroxymethyl derivatives:



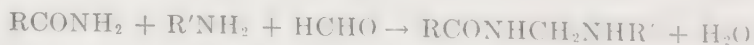
Positively charged amino groups do not undergo this reaction. The effect of adding formaldehyde is to shift the pH region, in which the amino groups are titrated, downward into a more acid range. In the presence of $0.3 M$ formaldehyde or more, the titration of the amino groups is essentially complete by pH 8.5, whereas in pure water it has scarcely begun at this pH . Thus, if the protein is titrated in water to pH 8.5, and formaldehyde is then added, the solution becomes more acid. The number of equivalents of OH^- ion which are then required to bring the solution back to pH 8.5 is a fairly accurate measure of the number of ϵ -amino groups in the protein.⁹ In β -lactoglobulin, this number is 33 to 35 groups per mole of protein, and the value of \bar{Z} , at the completion of this stage of the titration, is -52 to -53 . Under these conditions only the guanidinium groups of arginine residues should remain positively charged. The tyrosine hydroxyl groups would not begin to ionize appreciably until near pH 10; thus the only anionic groups at pH 8.5 in formaldehyde should be the ionized $-COO^-$ groups. Hence we may conclude that

$$(COOH) - (Arg) = 52 \text{ to } 53$$

The titration was not carried far beyond pH 10, so that no conclusions can be drawn from the work of Cannan *et al.* concerning the tyrosine hydroxyl groups of β -lactoglobulin.

Valuable further evidence concerning the nature of the groups ioniz-

⁹ The action of formaldehyde on amino acids, peptides, and proteins has been the subject of a vast literature. Beside reacting with amino groups, as described in the text, formaldehyde undergoes important reactions with sulfhydryl, amide, and other groups, and serves as a cross-linking agent by the formation of methylene-bridges, for instance between an amino and an amide group in different molecules, or in different parts of the same molecule:



A survey of the field has been given by French and Edsall (1945), and a more recent but briefer discussion by Putnam (1953). An important series of papers by Fraenkel-Conrat and his colleagues (see, for instance, Fraenkel-Conrat and Mechem, 1949)) should also be consulted.

ing at various stages of the titration is obtained by studying the effect of temperature on the titration curve. As was pointed out by Wyman (1939), one may interpret the change of pH with temperature, in a solution to which a fixed amount of strong acid or base has been added to the isoelectric protein, in terms of what may be called an apparent heat of ionization, Q' . This is defined by the relation

$$Q' = -2.303RT^2 \left(\frac{\partial pH}{\partial T} \right)_{\bar{n}} \quad (69)$$

In a simple buffer mixture of a monovalent acid and its conjugate base, in constant proportions, the change of pH with temperature would

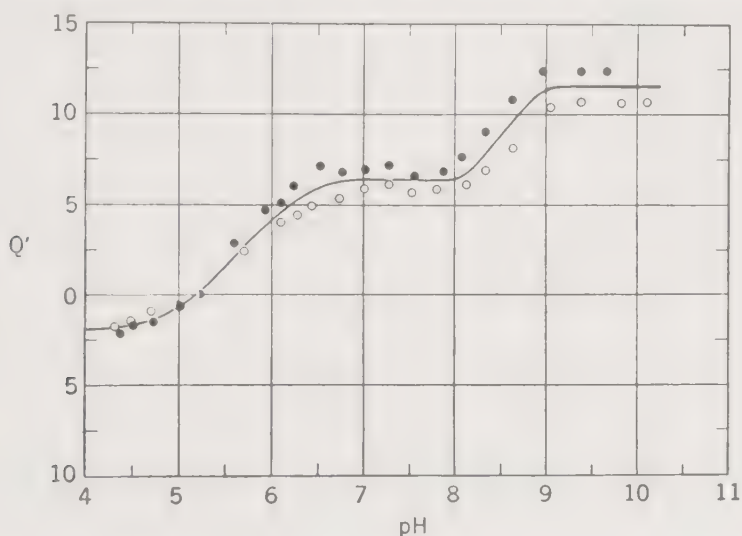


FIG. 10. Heat of ionization (Q') of the groups in horse hemoglobin as a function of pH . (From Wyman, 1939.)

obviously be related to the heat of ionization by (69)—compare Chapter 8, equation (79), and the general discussion of buffer action in that chapter. The transfer of this definition to a polyvalent acid is an obvious step. The results are perhaps seen with the greatest clarity in Wyman's titration of hemoglobin at three different temperatures between pH 4 and 10 (Fig. 10).

The low value of Q' , near -1 kcal/mole, is characteristic of carboxyl groups and is observed below pH 5. A higher value of Q' , 6 to 7 kcal/mole, is found from pH 6 to 8 and is similar to, though somewhat smaller than, the values found to be characteristic of imidazole groups in Table VI of Chapter 8. Finally, above pH 9, the value of Q' is much greater, near 11 kcal/mole; this value is apparent from the data of Chapter 8, Tables III and VI. There is a transition zone near pH 5.5, in which Q' is

changing from a value characteristic of carboxyl to that characteristic of imidazole groups, and another transition zone near pH 8.5, involving the transition from imidazole to amino groups. The three stages are particularly apparent in the case of hemoglobin, which has an unusually large number of imidazole groups (36 per mole in horse hemoglobin); but the same general sort of variation of Q' with pH is observable with other proteins. The data for β -lactoglobulin are plotted in Fig. 11, in which Q' has been plotted, not against pH , but against \bar{Z} . It is seen that an abrupt rise in Q' , from approximately 1 kcal/mole to 10 kcal/mole, occurs over

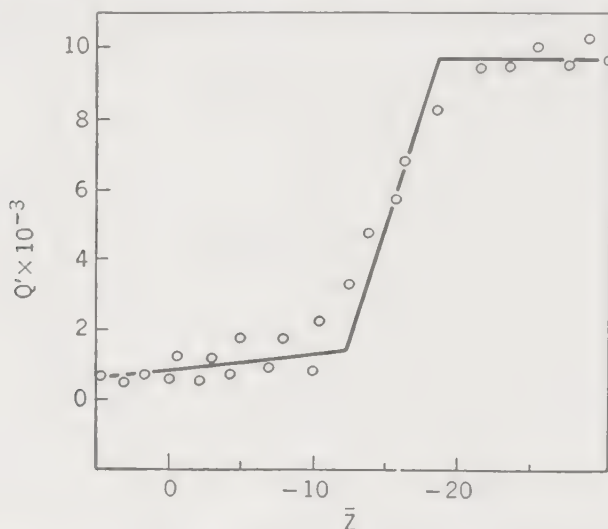


FIG. 11. Heat of ionization of the groups in β -lactoglobulin, as a function of the number of protons removed from the isoelectric molecule. (From Cannan, Palmer, and Kibrick, 1942.)

the range from $\bar{Z} = -12$ to $\bar{Z} = -18$. This suggests the presence of a relatively small number of imidazole groups, probably not more than 6, and very likely less, since there will be overlapping between the dissociation of imidazole groups and any α -amino groups that may be present as the terminal groups of peptide chains. The number of imidazole groups is so much smaller than in hemoglobin that there is no plateau region corresponding to a Q' value near 6 to 7 kcal/mole.

The reader should be warned against supposing that Q' for a given type of group in a protein molecule is *necessarily* nearly the same as for the same type of group in a simple peptide or other related small molecule. If certain acid or basic groups in the protein are involved in strong hydrogen bonds, or buried in the interior of the molecule instead of being on the surface, the values of Q' may be markedly altered thereby, as may their intrinsic pK values. We shall find actual instances of such behavior in several proteins.

We may now set up a balance sheet to compare the ionizable groups, as determined from the titration curve, with the numbers of such groups as calculated from the analyses reported in Chapter 3, Table II.

TABLE VII
IONIZABLE GROUPS IN β -LACTOGLOBULIN
(Assumed molecular weight 40,000)

Class of group	Titration data	Analytical data
(a) Total cations = (His) + (Lys) + (Arg) + (α -Amino)	46	44-45
(b) (COOH) - (Arg)	52-53	52
(c) (COOH) - (Arg) - (Lys)	18-19	20-21
(d) (His) + (α -Amino)	6	7
(e) (Lys) = $b - c$	33-35	31-32
(f) (Arg) = $a - d - e$	6	6
(g) (α -Amino)	—	3
(h) (His) = $d - g$	3	4
(i) (Asp)	—	34
(j) (Glu)	—	51
(k) (α -COOH)	—	3
(l) (Total CONH ₂)	—	30
(m) (COOH) = $b + f$ (titration) = $i + j + k - l$ (analysis)	58-59	58

Titration data from Cannan *et al.* (1942); analytical data from Tristram (1953), recalculated to molecular weight 40,000.

The number of α -carboxyl groups (3) is *assumed* to be the same as the number of free α -amino groups.

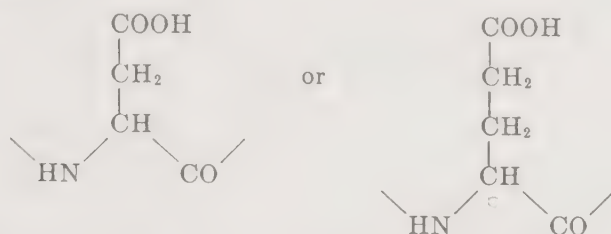
It is apparent that on the whole the number of groups of the various classes, calculated from the titration data, is in very satisfactory agreement with the number determined from amino acid analysis. There is a slight discrepancy between the two methods of estimating the number of lysyl residues, but this is not much greater than the uncertainty in the allocation of groups on the basis of the titration data. The titration method in this case cannot adequately distinguish between the histidyl residues and the α -amino groups, which are present in similar numbers and with almost identical pK values. The total for both kinds of groups, however, turns out to be about the same by either method. The number of carboxyl groups from analysis agrees remarkably well with that determined from titration; however, it should be remembered that the number of α -carboxyl groups is *assumed* to be three per molecule, and that

this assumption has not been experimentally proved as in the case of the number of α -amino groups.

Having assigned the number of groups of each class, it is possible to analyze the titration data at various ionic strengths (Fig. 9) to obtain a value for w , using equation (58). The analysis has been carried through in particular detail for the carboxyl groups, of which there are 58 in all. If α is the fraction of these groups in the ionized form at any pH value, then it is apparent from equation (58) that, if we plot $pH - \log [\alpha/(1 - \alpha)]$ against \bar{Z} , a straight line should be obtained. The slope of this line should be $-0.868w$, and the intercept on the ordinate axis when $\bar{Z} = 0$ gives pk_{int} . A deviation of the experimental data from linearity would indicate that some of the assumptions made in deriving (58) are not fulfilled. We note that \bar{Z} , which is identical with \bar{Z}_p in (68), should be calculated from the latter equation, taking account of *all* ions bound. For β -lactoglobulin in KCl solution, we can safely assume that no potassium ion is bound in neutral or acid solution. The work of Carr (1953) shows that no chloride ion is bound until the pH is somewhat acid to the isoelectric point. Thus $\bar{Z} = \bar{v}_{H^+}$ in the region near the isoelectric point and alkaline to it, but a small correction for chloride ion binding must be made in more acid solutions, the correction of course increasing with increase of (Cl^-) . The data for the necessary corrections can be obtained from Carr's work. Plots of the titration data for three different ionic strengths have been made by Tanford (1955) and are shown in Fig. 12. It is clear that (58) is indeed obeyed for this system, and that a unique value for w is thus obtained at each ionic strength. Moreover, the value of w varies with the ionic strength in the manner that would be predicted by (53). The radius, b , of the sphere equivalent to the β -lactoglobulin molecule, calculated from (53), is nearly the same for the three different ionic strengths shown and is close to 25 Å. We may compare this with the calculated dimensions of the molecule, derived from its molecular weight, M , and partial specific volume, v . If the molar volume is equal to $Mv = 40,000 \times 0.75 = 30,000$ ml, the molecular volume is Mv/N . This volume is also equal to $4\pi b^3/3$, if the molecule be considered as a sphere of radius b . This calculation gives $b = 23 \times 10^{-8}$ cm = 23 Å. If we assume that the molecule binds some solvent water, say 0.2 gram of water per gram of protein, and that the specific volume of the bound water is unity, the calculated b value would increase to 25 Å. The charged sphere model in this case is therefore a very plausible one, although we are not justified on this account in assuming that the actual protein molecule is truly spherical.

The value of pk_{int} for the carboxyl groups of β -lactoglobulin is seen from Fig. 12. to be 4.60 for all the three curves shown; all of them intersect at this point on the vertical line corresponding to $\bar{Z} = 0$ in the figure.

This is not an unreasonable value for an aspartyl or glutamyl residue in a peptide chain:



No pK values appear to be available for simple substances closely analogous to such residues. The monoesters of succinic acid, $\text{ROOC} \cdot \text{CH}_2\text{CH}_2\text{COOH}$, have pK values near 4.5, however, and the monoesters

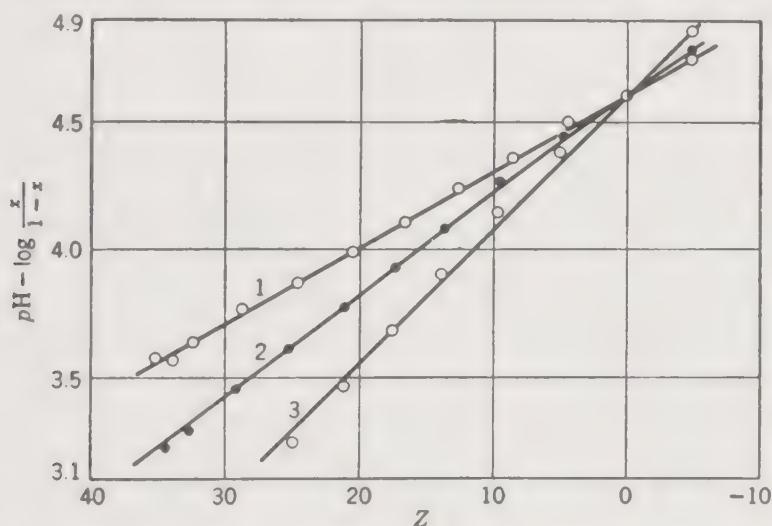


FIG. 12. The function $pH - \log [\alpha / (1 - \alpha)]$ as a function of the net charge Z on the β -lactoglobulin molecule. Curves 1, 2, and 3 are for ionic strengths 0.270, 0.069, and 0.019 respectively. (Data of Cannan, Kibrick, and Palmer (1942) as represented by Tanford (1955) with corrected values of net charge, taking account of chloride ion binding.)

of adipic acid, $\text{ROOC} \cdot (\text{CH}_2)_4 \text{COOH}$, near 4.6 (Cohn and Edsall, 1943, p. 121; and Tanford, 1955, p. 261). These may be regarded as moderately similar to glutamyl or aspartyl residues, in a peptide chain, and the agreement with the value of 4.6 for the carboxyl groups in the β -lactoglobulin therefore appears fairly plausible. It is surprising, however, that a single pK_{int} value is adequate to characterize all the 58 carboxyl groups of β -lactoglobulin, even though some of them are aspartyl and some are glutamyl residues.

Cannan *et al.* (1942) also obtained values of w from the portion of the titration curve involving the ionization of the histidine imidazole groups.

These values were quite consistent with those obtained from the ionization of the carboxyl groups, and the value of pK_{int} was found to be very close to 6.8 at 25°. This appears to be an entirely reasonable value for an imidazole group. Since the data of Table VII indicate, however, that the ionization of terminal α -amino groups is also involved in this pH range, the significance of these findings should not be too strongly stressed. The form of the more alkaline portions of the titration curve, at various ionic strengths, proved more difficult to interpret, and no constant value of pK_{int} could be obtained for these groups.

The analyses given in Chapter 3, Table II, indicate approximately 9 tyrosine residues for a molecular weight of 40,000, but the titration data of Cannan *et al.* do not give evidence as to whether these groups ionize in the pH range studied by them.

Ovalbumin

Ovalbumin, one of the best known of all proteins, has been intensively studied for more than two generations. Its general properties have been well summarized, with extensive references, in the reviews by Fevold (1950) and by Warner (1954) (see also Anfinsen and Redfield, 1956). Complete agreement on a molecular weight value has not been attained, but it appears to lie between 43,000 and 46,000. The molecule contains 3 or 4 sulfhydryl groups and 1 or 2 disulfide linkages. No free terminal amino group can be detected; this may be blocked by the presence of a carbohydrate group with a molecular weight of about 1200, which is known to be present. One C-terminal carboxyl group of proline has been detected. The amino acid composition of ovalbumin is listed in Chapter 3, Table II. A detailed study of the titration curve at various ionic strengths has been carried out by Cannan *et al.* (1941). Their titration curve covers the pH range from 2 to 11.5, and the analysis of the data which they carried out was similar to that which we have described in detail for β -lactoglobulin. The number of titratable groups of various sorts deduced from the titration curve is generally in good agreement with the values to be expected from the amino acid analyses, if allowance is made for the fact that ovalbumin contains one or two phosphoric acid residues, probably attached to serine or threonine residues.

One feature of the titration curves of ovalbumin is extremely revealing for the light it throws on the structure of this protein. Crammer and Neuberger (1943) studied its ultraviolet absorption, in the range from 250 to 320 m μ , as a function of pH. We have already seen (in Chapter 8, p. 426) that the ultraviolet light absorption due to the phenolic hydroxyl group of tyrosine changes markedly on ionization. The maximum molar extinction coefficient (ϵ) of tyrosine in acid solution is 1290 at a wave-

length of 275 $m\mu$. In strongly alkaline solution, the position of the maximum shifts to 295 $m\mu$, and the value of ϵ at the maximum increases to about 2300. The characteristic pK value of the phenolic hydroxyl group in tyrosine and simple tyrosyl peptides is close to 10 at 25°. This may be taken as the intrinsic value (pK_{int}) for such groups in proteins (see equation 58). Proteins with acid isoelectric points such as ovalbumin (pH_I near 4.6) carry a large negative net charge at pH values near 10. For electrostatic reasons this charge makes the ionization of the phenolic groups in the protein more difficult, and thereby increases the pH at which a given degree of ionization (α) of these groups is attained. This effect follows from equation (58), when \bar{Z} is negative. In a protein such as insulin, the transition from the absorption spectrum of un-ionized phenolic groups to that characteristic of almost completely ionized groups occurs over the pH range from 9.5 to about 12, as Crammer and Neuburger showed. Moreover, the spectrum at any pH up to 12 or thereabouts can be obtained reversibly, approaching the given pH value from either the acid or the alkaline side. The situation in ovalbumin is entirely different. As the pH is increased up to about pH 12, or even a little above, the absorption spectrum still remains characteristic of that of un-ionized phenolic hydroxyl groups. At pH values near 12.5 and above, there is a rapid and irreversible change; the phenolic groups ionize, as shown by the increase of absorption near and above 295 $m\mu$, and they remain ionized when the pH is brought back to 12 or a little below. Evidently in the native protein the phenolic hydroxyl groups are somehow bound or blocked so that they are not free to ionize. It seems probable that this is due to some form of hydrogen bonding between them and other groups in the molecule—a bond which is broken only at very high pH values.

An extremely simple model for this kind of hydrogen bond is shown by salicylic acid (Chapter 8, Table VI) in which the pK value of the phenolic group is displaced from its usual value near 10 to approximately 13. Obviously in this case the hydrogen of the hydroxyl group is held by a strong hydrogen bond to the neighboring ionized carboxyl group in the *ortho* position. No proof has yet been given that the bonds which involve the phenolic hydroxyl groups of ovalbumin also involve ionized carboxyl groups of glutamic or aspartic residues, but this is one possible interpretation of the experimental findings. In any case, the results indicate that the tyrosyl residues of the ovalbumin molecule in its natural undenatured state are unavailable for reaction with bases, and become available only after a rather drastic change in the internal structure of the molecule. It seems that the tyrosyl groups serve to stabilize the native structure of the molecule and hold it in a rather close-knit configuration. The native structure can be broken down, and the tyrosyl groups made available for

reaction with bases, as Crammer and Neuberger showed, by denaturation in alkali, and also by denaturation in acid, in concentrated urea solutions, or by heat. This particular denaturation process appears to be quite irreversible.

These experiments indicate that the amino acid analysis of a protein is not an adequate guide to the character of its titration curve. We must always be prepared for the possibility that certain groups may be somehow bound into the native structure by hydrogen bonding, or made inaccessible for reaction by being located in the interior of the molecule, where they cannot be attacked directly by the solvent. This is analogous to the unreactivity of many sulfhydryl groups in native proteins with reagents specific for such groups, whereas such sulfhydryl groups often become readily reactive after the protein is denatured (see Chapter 3).

Serum Albumins

Human and bovine serum albumins have been studied perhaps more intensively than any other proteins by the physical chemists. Serum albumin is the major protein constituent of blood plasma, with a molecular weight which has been variously estimated from 65,000 to 69,000. It contains approximately 100 ionizable carboxyl groups, and a nearly equal number of groups which are positively charged at the isoelectric point. We have already given some discussion of the properties of the serum albumins, and they are well discussed in a more detailed fashion by Hughes (1954, pp. 677–698).

The titration curve of bovine serum albumin has been studied in detail by Tanford *et al.* (1955c), who assumed a molecular weight of 65,000. The appearance of the curve at four different ionic strengths is shown in Fig. 13. Qualitatively the effect of varying ionic strength is similar to that on other proteins; the curves grow steeper, around the isoionic point, with increasing ionic strength. Heats of ionization, determined from the change of the titration curve with temperature, indicated that the first 100 groups titrated, from pH 2 to about 5.5, gave values of the apparent heat of ionization, Q' (equation 69), not far from zero, as would be expected for carboxyl groups. For the imidazole groups, ionizing in the region from pH 5.5 to 8.5, Q' was near 6.5 kcal/mole, and for the amino groups above 10 kcal/mole. The state of ionization of the tyrosyl groups could be inferred from the change of ultraviolet absorption with pH; such data had already been obtained by Tanford and Roberts (1952). From all these data the number of groups of the various kinds, and their intrinsic pK values, could be deduced and compared with the values inferred from amino acid analysis (Table VIII). The results indicate that serum albumin deviates from most other proteins in the values found for

pK_{int} for the carboxyl, ϵ -amino, and phenolic hydroxyl groups. The values for the carboxyl and ϵ -amino groups are unusually low, and those for the phenolic hydroxyl groups distinctly higher than for insulin or ribonuclease, or for simple tyrosyl peptides, although they are lower than for lysozyme, which behaves in a rather abnormal fashion. Moreover the heats of ionization of the phenolic hydroxyl groups in serum albumin are unusually high, with $\Delta H^\circ = 11.5$ kcal/mole, instead of the usual value

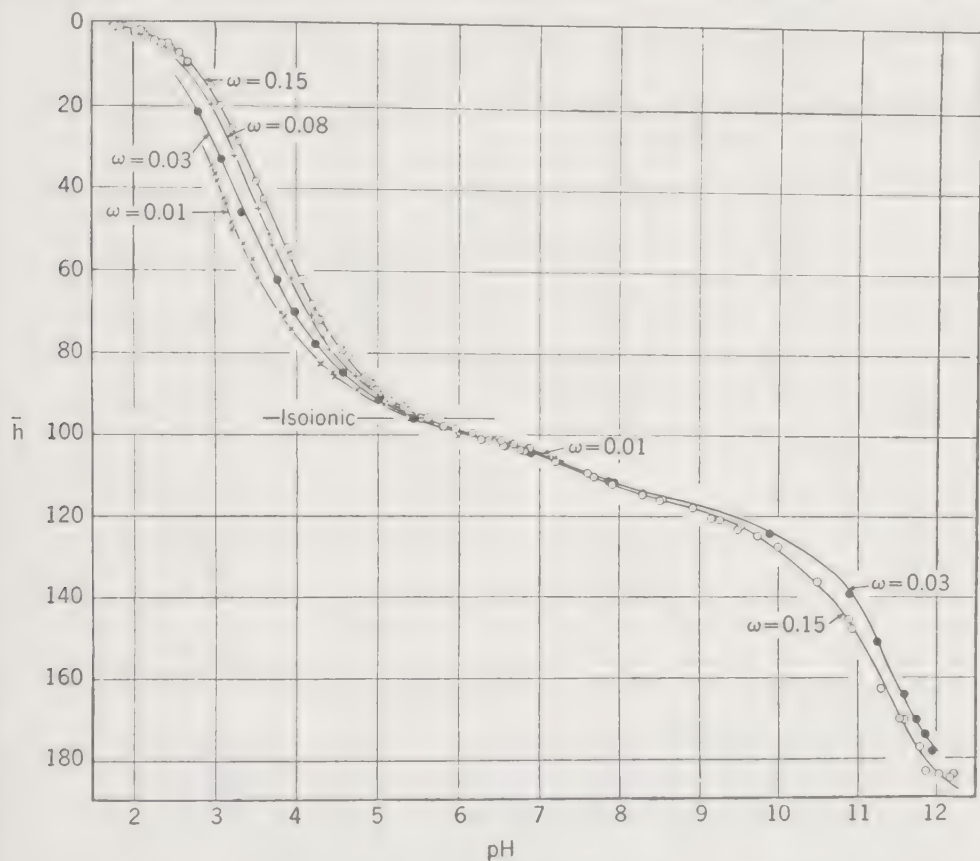
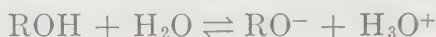


FIG. 13. Titration curve of bovine serum albumin at three different ionic strengths. (From Tanford, Swanson, and Shore, 1955.)

of 6.0 for phenolic hydroxyl groups (compare data in Chapter 8, Table VI). Correspondingly the standard entropy of ionization, ΔS° , is only -9 cal deg $^{-1}$ mole $^{-1}$ for these groups in serum albumin, instead of the typical value of -22 to -25 for phenolic hydroxyl groups. All this suggests that these groups are involved in some kind of labile linkage within the molecule which has to be broken before the phenolic $-\text{OH}$ group is free to ionize. The breakage of these links, if it freed tyrosyl and perhaps other side chains from constraints which inhibit their freedom of motion in the native molecule, would contribute a positive entropy term par-

tially counterbalancing the large negative entropy charge associated with the orientation of water molecules around the charged groups formed by the reaction



Qualitatively this might explain the numerically low ΔS° value which is observed. In any case the situation is entirely different from that in ovalbumin, for the phenolic hydroxyls in serum albumin ionize readily and reversibly at $p\text{H}$ values between 10 and 12; hence any linkages which are broken before these groups ionize are re-formed when the solution is brought back to a lower $p\text{H}$.

TABLE VIII
IONIZABLE GROUPS IN BOVINE SERUM ALBUMIN AND THEIR INTRINSIC $p\text{k}$ VALUES
(Assumed molecular weight of albumin 65,000)

Group	Number per molecule, titration	Number per molecule, amino acid analysis*	Albumin $p\text{k}_{\text{int}}$	Other proteins† $p\text{k}_{\text{int}}$
α -Carboxyl	(1)	1	(3.75)	3.6
Carboxyl (Asp and Glu)	99	101	3.95‡	4.3–4.7
Imidazole (His)	16	17	7.0	6.4–7.0
α -Amino	(1)	1	7.8	7.4–7.9
ϵ -Amino (Lys)	57	57	9.8	10.1–10.6
Phenolic (Tyr)	19	18	10.35	8.5–10.9
Guanidine (Arg)	22	22	>12	11.9–13.3
Sulphydryl (CySH)	(0)	(0.7)	—	—

Values taken from the study by Tanford *et al.*, 1955c.

* See Chapter 3, Table II.

† The other proteins, for which comparison data are given here, include insulin (Tanford and Epstein, 1954), β -lactoglobulin (Cannan *et al.*, 1942), ovalbumin (Cannan *et al.*, 1941), lysozyme (Tanford and Wagner, 1954), and ribonuclease (Tanford *et al.*, 1955b; Tanford and Hauenstein, 1956a). The $p\text{k}_{\text{int}}$ values for the phenolic groups differ markedly for different proteins, being 9.7 ± 0.2 for insulin and ribonuclease, and 10.6 ± 0.3 for lysozyme. Three of the phenolic groups in ribonuclease are totally unreactive in the native molecule and can be titrated only after denaturation.

‡ The $p\text{k}_{\text{int}}$ value for the carboxyl groups varied with ionic strength, from 3.92 at $\omega = 0.01$, to 4.02 at $\omega = 0.15$.

The low $p\text{k}_{\text{int}}$ values for the carboxyl and ϵ -amino groups suggest that these groups also are somehow bonded, or modified by interaction with other neighboring groups, in the native molecule. In our present state of ignorance concerning the detailed molecular structure of albumin, however, it seems useless to speculate on the nature of these interactions.

Tanford *et al.* (1955c) evaluated the parameter, w (equations 53 and

58), from their titration data. They calculated the net charge, \bar{Z} (equation 58), which must be known to calculate w , by combining their own data for the binding of protons with those of Scatchard *et al.* (1950) for the binding of chloride ions; the cations (K^+ ions), had been found by these authors to be unbound. Using the type of procedure involved in equations (65) to (68) above, Tanford *et al.* obtained values of w which were independent of pH over the range 4.2 to 10.3; the variation of w with ionic strength was consistent with the values calculated from (53),

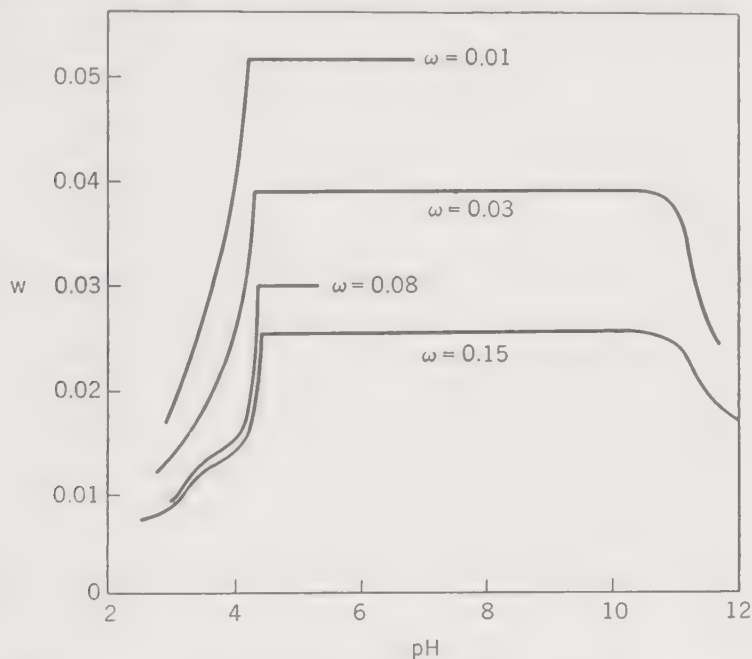


FIG. 14. The variation of the value of w with pH at different ionic strengths, for bovine serum albumin. (From Tanford, Swanson, and Shore, 1955.)

if serum albumin is represented by a spherical model of radius $b = 33$ Å approximately. This is somewhat larger than the value of 30 Å calculated on p. 512 for a sphere with 20% hydration, but considering the limitations of the model the agreement seems very reasonable.

The value of w , however, decreases sharply at pH values acid to about 4.2, and also at strongly alkaline pH values. The variation with pH , at different ionic strengths, is shown in Fig. 14. If we assume that a spherical model is still applicable at low and high pH values, the decrease of w implies an increase in the radius of the sphere (equation 53). That is, the molecule expands as its net charge increases, although on this interpretation the expansion sets in on the acid side at a relatively low positive value of \bar{Z} of +5 to +10, whereas on the alkaline side expansion does not occur until \bar{Z} is greater than -50. In any case, the decrease of

w indicates that some important reversible intramolecular alteration is going on in the albumin molecule, in acid solutions. Evidence that this change involves an expansion of the molecule has also been obtained from viscosity and other types of measurement (Yang and Foster, 1954; Tanford and Buzzell, 1956; Tanford *et al.*, 1955a). Different interpretations of the phenomena have been advanced by Loeb and Scheraga (1956) and have been criticized (Foster and Aoki, 1957), but the discussion of all these problems is outside the scope of this chapter. These observations show, however, that the study of the titration curves of proteins may furnish important information concerning changes in molecular shape, or in the internal bonds holding different parts of the molecule together, although the complete interpretation of such data is not yet achieved.

Ribonuclease

Since the amino acid composition of beef pancreas ribonuclease is now completely known (Chapter 3, Table II), and the sequence of amino acid residues in the single peptide chain of 124 residues has been largely worked out (Chapter 3, Fig. 6), the study of its titration curve is of particular interest. The measurements of Tanford and Hauenstein (1956a) show complete accord between the titration curve and the values to be expected from amino acid analysis, as indicated in Table IX.

TABLE IX
RIBONUCLEASE: COMPARISON OF TITRATION DATA AND AMINO ACID ANALYSIS

Group	Found by titration	Predicted from amino acid content	Intrinsic <i>p</i> k	
			Observed	Usual range of values
α -Carboxyl	(1)	1	(3.75)	3.75
β, γ -Carboxyl* (Asp and Glu)	10.2	10.2	4.0; 4.7	4.6
Imidazole (His)	4	4	6.5	6.5-7.0
α -Amino	(1)	1	7.8	7.8
ϵ -Amino (Lys)	10	10	10.2	10.1-10.6
Phenolic (Tyr)†	3 } 3 }	6	{ 9.95 Inaccessible	9.6
Guanidyl (Arg)	4	4	>12	>12

Tanford and Hauenstein, 1956a.

Values in parentheses were assumed.

* The "predicted" value of 10.2 is not an integer, because the preparation studied was a mixture of ribonuclease A and B, which differ by a single free carboxyl group, but otherwise appear to be identical in amino acid composition and structure (Tanford and Hauenstein, 1956b).

† The "inaccessible" phenolic groups are discussed in the text.

Because of the excess of free basic over carboxyl groups, the isoelectric point ($\bar{Z} = 0$ at pH 9.7) is more alkaline than for the other proteins we have considered. Light scattering and viscosity measurements gave no indication of change of molecular weight or shape between pH 2 and 11. It therefore is to be expected that w (equations 53 and 58) should remain constant over this entire range. For the carboxyl groups, however, it proved impossible to fit the data with a single value of pk_{int} , using the values of w derived from the ionization of the three reversibly reacting phenolic groups. The best explanation appeared to be that the carboxyl groups could be divided into two classes, one with $pk_{int} = 4.0$, the other with $pk_{int} = 4.7$, there being approximately five groups in each class. Those with the more acid pk_{int} values might behave as they do because of being located near positively charged groups; and the structure shown in Chapter 3, Fig. 6, indicates that several free carboxyl groups in ribonuclease do lie near such positive charges—for instance, the glutamyl residue at position 2, adjoining the N-terminal lysyl residue with its two positive charges, and the aspartyl residue at position 38, which lies between a lysyl and an arginyl residue. Two or three other carboxyl groups in the sequence may be near positively charged groups, in regions where the sequence is not yet completely known. Five carboxyl groups, on the other hand, definitely cannot lie near positively charged groups, unless through folding of the peptide chain. These considerations, pointed out by Tanford and Hauenstein, indicate the possibility of detailed correlations between titration curves and structure which should become increasingly frequent as structures become better known.

The six phenolic groups of the tyrosyl residues in ribonuclease fall into two sharply distinct classes. Three ionize readily and reversibly, with $pk_{int} = 9.95$ (Table IX). The other three are inaccessible to titration in the native protein, like the phenolic groups of ovalbumin; they are gradually released on standing in alkaline solution, at pH values above 12, and then remain reactive with hydrogen ions at lower pH values. The process of release is irreversible and is accompanied by loss of enzymatic activity (Shugar, 1952; Tanford *et al.*, 1955b). It will be of great interest to know just which tyrosyl residues are inaccessible in the native molecule, and to determine how they are made so in the native structure. This problem should soon be accessible to experimental attack.

Hemoglobin

The mammalian hemoglobins are notable for their high content of histidyl residues—36 per mole in horse hemoglobin, out of a total of about 540 residues—and consequently are excellent buffers in the phys-

iological pH range, since the pK_{int} values of the imidazole groups lie near pH 6.5 to 7.0. The heat of ionization of horse hemoglobin at various pH values has already been shown in Fig. 10. Here we comment only on one special feature of the titration curve of hemoglobin—the release of a large number of titratable groups, on exposure to acid, which were inaccessible in the native protein. This phenomenon has been studied in great detail by Steinhardt and Zaiser (1951, 1953, 1955). They mixed horse carboxyhemoglobin and acid rapidly in a mixing chamber, from which the solution flowed through the pH meter, so that pH determinations could be made on solutions which had been mixed as little as 3 seconds

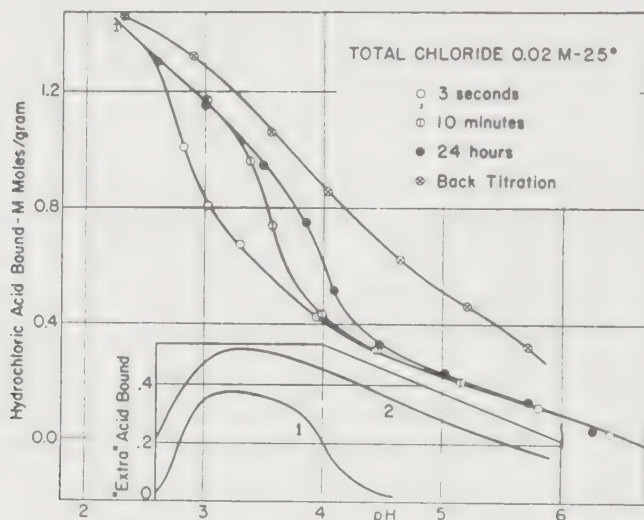


FIG. 15. The binding of protons (from HCl) by horse carboxyhemoglobin at constant chloride ion concentration, as a function of pH and time. The two curves in the inset are obtained by difference: curve 1, by subtracting the 3-second data from the 24-hour data; curve 2, by subtracting the 3-second data from the back-titration data. (From Steinhardt and Zaiser, 1955.)

beforehand. Such freshly mixed solutions bound much less hydrogen ion below pH 4.5 than solutions which had been allowed to stand several minutes after mixing. Some of the different acid titration curves obtained after different time intervals are shown in Fig. 15. The speed with which the inaccessible groups are released increases rapidly as the pH is lowered; below pH 3 it becomes so fast that the change is largely complete in considerably less than 3 seconds. Between pH 3 and 6.6, the 3-second data indicate the titration curve of the native protein; the back-titration data of Fig. 15 represent the titration curve of protein in which the previously unreactive groups have been liberated. The difference, shown by curve 2 of the inset in Fig. 15, thus gives the number of groups per mole released by the acid treatment. The maximum difference,

0.52 mM/g, or 35 to 36 groups per mole of hemoglobin, is obtained between pH 3.1 and 3.5.

The same release of 36 groups which were not titratable in the native protein occurs if ferrihemoglobin (methemoglobin) is studied instead of carboxyhemoglobin. In ferrihemoglobin, however, the process appears to be fully reversible when the solution is returned to a higher pH value near neutrality, whereas with carboxyhemoglobin the reversibility is only partial.¹⁰ Thus at any given pH there is an equilibrium between the native hemoglobin and a denatured form in which 36 extra groups can be titrated in acid solution. Above pH 4.5 the native form is overwhelmingly predominant at equilibrium; below pH 3.6 the denatured form. The rate of transformation can be followed not only by titration measurements (Fig. 15) but also by spectroscopic changes (light absorption at 405 $m\mu$) and by determining the amount of protein precipitated on bringing the solution rapidly to pH 7, since the denatured protein is very insoluble near the isoelectric point. All these methods give good agreement in determining the proportion of protein in the denatured state at any time. This is important, since it indicates that the process of releasing the titratable groups occurs in an "all-or-none" fashion; a molecule either has all 36 extra groups available for reaction, or none of them; the transition period between these two limits must be extremely brief. These extra groups, in view of the pH range in which they are titratable, must be for the most part, and perhaps entirely, carboxyl groups; a few might be imidazole groups of histidine. The titration curves of Wyman (1939) show, however, that nearly all the imidazole groups are already titratable in the native protein. Steinhardt and Zaiser (1955), however, have concluded that the ϵ -amino groups of lysine are probably in large part unreactive in the native protein. Indeed, some such arrangement appears necessary if we consider the amino acid composition of hemoglobin (Chapter 3, Table II) and the fact that its isoelectric point is near pH 7 in both the native and the denatured form. If anionic groups ($-\text{COO}^-$) are blocked so as to be unreactive in the native form, and released on denaturation, then the same must be true of an approximately equal number of cationic groups (ϵ -amino or possibly guanidino), if the isoelectric point is to remain nearly the same. The whole problem is discussed critically and in great detail by Steinhardt and Zaiser (1955).*

¹⁰ If ferrihemoglobin is allowed to stand for some time in acid solution, slow, irreversible changes occur; but for experiments of short duration these irreversible changes can be neglected.

* Tanford (1957a) has recently shown that the viscosity increment of hemoglobin in acid solution, under the conditions described by Steinhardt and Zaiser, is much greater than for native hemoglobin in neutral solution. This increase in viscosity may

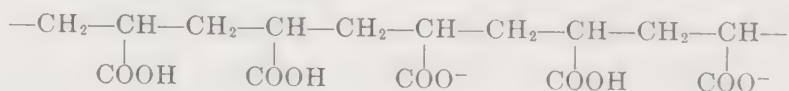
Hemoglobin is notable for its possession of "heme-linked acid groups"—that is, groups which change in acid strength when oxygen (or carbon monoxide) is bound to, or removed from, an iron atom on one of the four heme groups. It is a necessary consequence that the oxygen affinity of the heme iron is a function of the state of ionization of the neighboring heme-linked acid group. This set of phenomena, commonly referred to as the Bohr effect,¹¹ is of profound physiological importance (Henderson, 1928). The nature of the underlying mechanisms has previously been discussed in detail by one of us (Wyman, 1948), and we shall consider it again in the second volume of this book.

Other Proteins

No attempt will be made here to give a systematic survey of data on other proteins. We may call attention to the interesting titration studies of Tanford and Epstein (1954) on insulin and those of Fromageot and Schnek (1950) and of Tanford and Wagner (1954) on lysozyme.

Synthetic Polyelectrolytes, Including Polypeptides

In recent years many synthetic polymers containing acidic or basic side chains have been prepared. A typical simple example is polyacrylic acid, which when partially ionized may be described by such a formula as



Here the monomer unit is $(\text{—CH}_2\text{—CH—COOH})$.

In a chain structure made up of many such units, it is obviously reasonable to regard all the acidic or basic groups as intrinsically equivalent,

be inferred to indicate an expansion of the molecule. (See the discussion of viscosity in Vol. II.) Moreover, Field and O'Brien (1955) have shown that hemoglobin dissociates into half molecules under the same conditions. The combined dissociation and expansion of the molecule should result in a marked decrease in electrostatic interactions between the charged groups, and this should permit a larger number of protons to be bound at a given *pH*. (Compare equations (53) and (58).) Tanford estimates that this electrostatic effect may account for most, and perhaps nearly all, of the increased proton binding of denatured hemoglobin. Further experiments are required to determine whether this interpretation is adequate, or whether it is necessary to assume a more specific blocking mechanism, making some of the groups unavailable in the native molecule.

¹¹ The name is derived from that of the eminent Danish physiologist Christian Bohr, who in 1904 discovered the effect of varying carbon dioxide pressure on the oxygen affinity of hemoglobin. Later it was shown that acids other than carbonic acid also produced the same effects.

except for a few at the extreme ends of the chain which are relatively negligible if the chain is long. Obviously there will be strong electrostatic interactions between the charged groups in the chain; these will give rise to a spreading of the titration curve over a wider region of pH than would be true in the absence of such interactions. Qualitatively the result is similar to that produced by the term involving w in equation (58) and represented graphically in Fig. 8. There will in general be some binding of the counter-ions present in the solution to balance the charge on the polyelectrolyte; in the case of the polyacrylate ion these are generally sodium or potassium ions, but other counter-ions may of course be added also, and if they are divalent or trivalent they will generally be bound more strongly than the ions of the alkali metals. Other polyelectrolytes may contain amino groups or other uncharged basic groups in the repeating monomer unit; still others have been prepared which are amphoteric.

If the fraction, α , of all the acidic groups in a polyelectrolyte is ionized,¹² the distribution of ionized and un-ionized groups along the chain will not be completely random; placing a negative charge on any given $-\text{COO}^-$ group makes it less probable that there will be a negative charge on one of its nearest neighbors, for obvious electrostatic reasons.¹³ On the other hand, consideration of the entropy of the various possible distributions of charged and uncharged sites would favor a random distribution of the charges among the sites (compare Chapter 4, p. 229). The actual distribution must represent a compromise between these conflicting energy and entropy requirements.

Finally, a polyelectrolyte chain is a flexible structure which can assume a huge variety of different configurations, from a highly compact coil to an almost completely extended rodlike structure, by rotations around the valence bonds in the chain. Increasing the net charge obviously favors the more extended configurations, so that the geometry of the molecule is a function of α . We have encountered a situation of this sort in the expansion of the serum albumin molecule which appears to

¹² If the polyelectrolyte is a polyamine, or in general consists of monomer units which are uncharged in the basic form and positively charged in the form of the conjugate acid, the fraction of ionic groups is $1 - \alpha$, in accord with the notation we have used in this chapter.

¹³ This of course is true not merely for a chain structure but for any model in which the groups which may carry charges are in specified positions and are close enough to interact appreciably. In the sphere model used in this chapter we have avoided this difficulty by assuming the charge to be always spread uniformly over the surface of the sphere. This assumption enormously simplifies the calculations but is obviously not realistic in terms of a detailed structural picture. Nevertheless, as the evidence given in this chapter shows, it works surprisingly well for describing the titration curves of many proteins and polypeptides.

occur acid to its isoelectric point, and also apparently in strongly alkaline solution. In serum albumin, however, the polypeptide chain is cross-linked by 17 disulfide linkages, so that the expansion which can take place is severely restricted; whereas with an open-chain molecule, such as polyacrylic acid, freedom to stretch out is far greater. Indeed if such long-chain structures are drawn out into fibers, the fiber extends in a medium in which it acquires a large net charge, and contracts if the charge is neutralized. This furnishes a means of transforming chemical energy into mechanical work, which has been studied experimentally and theoretically by a number of investigators, notably by A. Katchalsky and his associates (Katchalsky, 1954a; Katchalsky and Zwick, 1955).¹⁴

A detailed discussion of the problems of acid-base equilibria in polyelectrolytes is beyond the scope of this chapter; we refer the reader to some of the important papers dealing with this subject, in which extensive references to other work will be found (Hermans and Overbeek, 1948; Overbeek, 1948; Pals and Hermans, 1952; Alfrey *et al.*, 1951; Katchalsky and Lifson, 1953; Lifson and Katchalsky, 1954; Katchalsky and Miller, 1954; Katchalsky, 1954b; Katchalsky *et al.*, 1954; Harris and Rice, 1955, 1956; Rice and Harris, 1955, 1956; Fuoss *et al.*, 1951).

One class of synthetic polyelectrolytes, however, is of particular biochemical interest—namely the polypeptides, or polyamino acids (Katchalski, 1950; Bamford *et al.*, 1956). These may be polymers of a single amino acid, such as poly-L-alanine, poly-L-lysine, poly-L-aspartic acid, poly-L-tyrosine; they may be copolymers of two or more amino acids, such as tyrosine and lysine, or alanine and aspartic acid, in almost any proportion. They may be multichain polyamino acids, such as those recently synthesized by Sela *et al.* (1956), which contain a series of peptide side chains, each branching from a functional side-chain group of a primary peptide chain—for instance multipoly-L-glutamyl-poly-L-lysine, in which each ϵ -amino group of the poly-L-lysine chain serves as the starting point for a peptide chain of glutamyl residues.

Polyamino acids with acidic or basic side chains have been studied particularly by E. Katchalski and his collaborators. A simple example of the results is shown in Fig. 16, which shows the ionization of poly-L-tyrosine as a function of pH , at ionic strength 0.2 in sodium chloride, and as extrapolated to zero ionic strength. The molecules in the preparation studied contained on the average 30 tyrosyl residues in the peptide chain (molecular weight near 5000). The fraction, α , of the phenolic hydroxyl

¹⁴ It has naturally been proposed that such a mechanism may explain the process of muscular contraction, but there are strong reasons for believing that muscular contraction operates by quite a different mechanism (see, for instance, Weber and Portzehl, 1952, 1954; Weber, 1957).

groups ionized was determined by spectrophotometry at $242\text{ m}\mu$ and at $293.5\text{ m}\mu$; the same values of α were derived from measurements at either wavelength. The curves drawn in Fig. 16 were calculated from (58), taking $pK_{\text{int}} = 9.5$, and $w = 0.0658$ at ionic strength 0.20. The net charge, \bar{Z} , in (58) was taken equal to 30α , since there are 30 groups per molecule, and binding of counter-ions was neglected. In calculating w from (53), the radius b was taken as 17.5 \AA , and a as 19.5 \AA , these values being calculated from the molecular weight of 5000. The agreement of the data

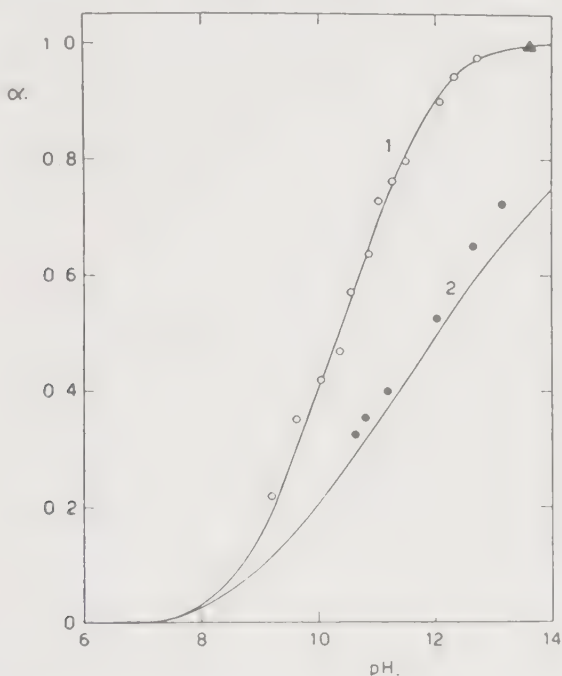


FIG. 16. Spectrophotometric titrations of poly-L-tyrosine. Open circles, experimental points at ionic strength 0.2 (NaCl). Solid circles, experimental values obtained by direct titration with sodium hydroxide in absence of salt, extrapolated to zero ionic strength. The point denoted by the solid triangle was obtained at $pH\ 13.6$. Curves 1 and 2 were computed from equation (58); see text. (From E. Katchalski and M. Sela (1953), *J. Am. Chem. Soc.* **75**, 5284.)

with the calculations for a spherical model is remarkable and suggests that the preferred shape of the peptide is a coil which is not far from spherical. Study of the effect of ionic strength on α at a constant pH (11.5) also yielded excellent agreement with the predicted changes from (58), the value of α at this pH rising from slightly above 0.4 at zero ionic strength to 0.8 at 0.2 ionic strength. Moreover, the value of pK_{int} obtained (9.5) was in close agreement with that (9.6) deduced from other measurements on tyrosine derivatives by Tanford and Roberts (1952). The titration curve was completely reversible on addition of acid. All this indi-

cates that the phenolic hydroxyl groups in poly-L-tyrosine are free to ionize and are not involved in any internal restrictions, such as those found in native ovalbumin or ribonuclease. Somewhat similar results were found by Katchalski and Sela for polydiiodo-L-tyrosine, except that pK_{int} was 7.7, a value much lower than for poly-L-tyrosine, but significantly

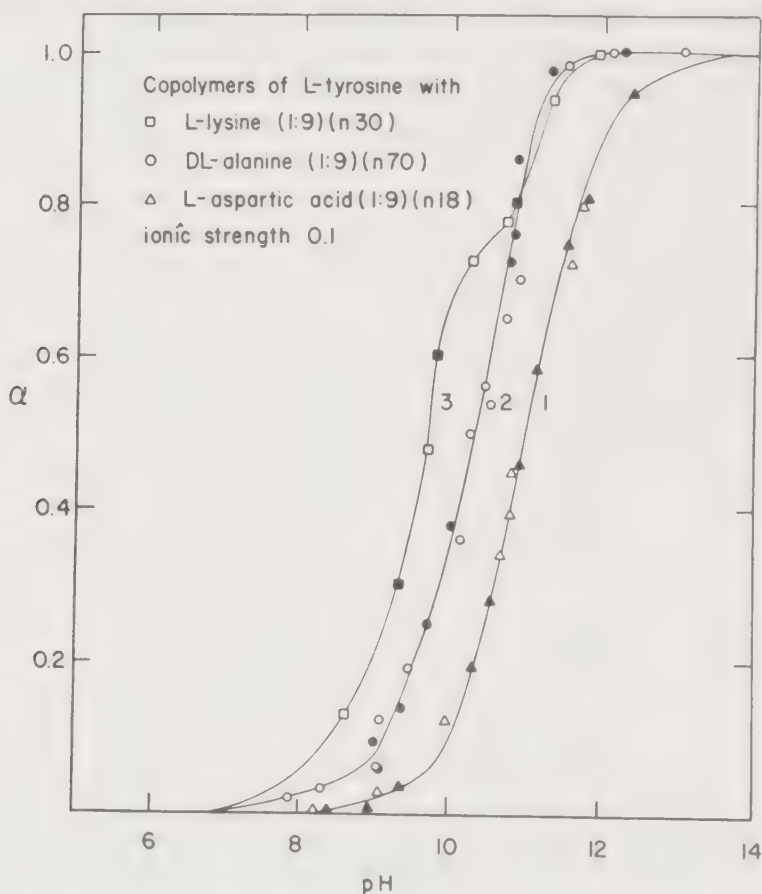


FIG. 17. Spectrophotometric titrations of copolymers of L-tyrosine with L-lysine, DL-alanine, and L-aspartic acid. The ratio of tyrosyl residues to those of the other amino acid in the copolymer was 1 to 9 in all cases. The symbol \bar{n} denotes the average number of residues per polymer molecule. The solid points represent back-titrations, i.e. results obtained after previous exposure of the copolymers to 0.1 N sodium hydroxide. (From M. Sela and E. Katchalski (1956), *J. Am. Chem. Soc.* **78**, 3986.)

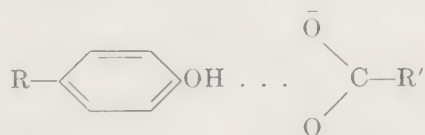
higher than that of 6.48 which has been obtained for monomeric diiodo-L-tyrosine. This discrepancy remains unexplained at present.

More recently, Sela and Katchalski (1956) have studied copolymers of tyrosine with alanine, lysine, and aspartic acid. Some of the titration curves of such copolymers are shown in Fig. 17. These measurements furnish a striking experimental confirmation of the prediction (see equations 58 to 60 and Fig. 8) that the titration curve for any given set of groups

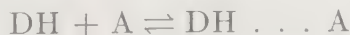
will be displaced to lower pH values by the presence of other positively charged groups in the molecule, and to higher pH values by the presence of other negatively charged groups. Curve 1 for the tyrosyl-aspartic acid copolymer in Fig. 17 could be well fitted by equation (58), with pK_{int} again equal to 9.5 and a radius, b , of 15.4 Å. Curve 2 for the tyrosine-alanine copolymer could not be fitted with a single value of w ; instead w decreased with increasing α , suggesting molecular expansion as the net charge increased. The tyrosine-lysine copolymer (curve 3), unlike the other two, is amphoteric, the lysyl residues giving up their positive charge and the phenolic groups acquiring a negative charge in the same pH region. The titration curve shows a jump near pH 10.6; similar jumps occur with other amphoteric polyelectrolytes at the isoelectric point (Katchalsky and Miller, 1954). It seems likely that this is associated with a change of configuration, the molecule coiling up rather tightly at the isoelectric point, and expanding on either side of this point when it acquires a net charge.

Effects of Internal Hydrogen Bonding on Titration Curves

Laskowski and Scheraga (1954) have developed a systematic treatment of the effects of internal hydrogen bonding, of various types, in order to evaluate their possible effects on observed titration curves. They classify several types of such bonds; for instance, a heterologous single bond is exemplified by a phenolic-carboxylate type of bond such as exists in the salicylate ion and may exist in proteins:

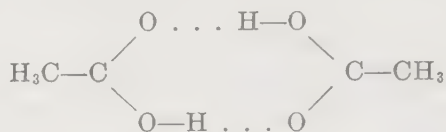


Here the phenolic group is the hydrogen donor (DH), the carboxylate group the acceptor (A), and the hydrogen-bonding equilibrium may be written



The proton might conceivably be transferred to A in the complex to form $\text{D} \cdots \text{HA}$, although in the particular case under discussion here this is most unlikely. Moreover the free DH group may ionize to give the conjugate base, D, and the free A group may acquire a proton and be converted into the conjugate acid, HA. All the different equilibria involved are considered by Laskowski and Scheraga, and their effects on the observed titration curves systematically formulated. They consider also homologous single hydrogen bonds—for instance $\text{NH} \cdots \text{N}$ bonds

between two imidazole groups—and homologous double bonds such as those in the dimer of acetic acid:



They also considered a variety of slightly more complicated possibilities. They showed that different types of hydrogen bonding could produce steepening or flattening of titration curves of a sort which has been generally attributed to the inaccessibility of certain reactive groups, or to unfolding of a compact macromolecule, such as a native protein. It is as yet extremely difficult to apply these general concepts to specific titration curve data, but the possibility of such applications will increase as knowledge of the details of protein structure increases.

Calculations with a Dielectric Sphere Model Containing Fixed Charges

We have already pointed out that a sphere with the net charge \bar{Z} spread uniformly over its surface is an unrealistic model in many ways for a macromolecule such as a globular protein. A decidedly more realistic spherical model has recently been proposed by Tanford and Kirkwood (1957). The sphere, of radius b , is treated as a small piece of dielectric, characterized by its dielectric constant D_i , and immersed in a continuous medium of dielectric constant D where $D \gg D_i$. The sphere is assumed to be impenetrable to the solvent and to contain a specified number of actually or potentially charged sites, in specified positions, either on its surface or within it. These sites are of two major types: cationic sites, which may carry a charge of $+\epsilon$ or zero, and anionic sites, which may carry a charge of $-\epsilon$ or zero. The sites of each type may fall into several classes, each characterized by an intrinsic standard free energy for proton binding, when one site of the given class binds a proton with all other sites discharged. Starting with all sites discharged, one may then calculate by complicated but straightforward electrostatic equations the work of placing charges on any specified set of sites, either in the absence of an ion atmosphere or at a specified finite ionic strength. Given the knowledge of this work of charging, for all possible distributions of charges among the sites, the form of the titration curve is determined. The calculations are naturally more complicated than for a uniformly charged sphere, and require a specific assignment of the positions of the acid and basic sites, on or within the sphere of radius b . Tanford (1957b) has made calculations for a number of simplified models, with various types of distribution of the sites. Although the models chosen

do not correspond, so far as is known, to any biological macromolecule, the results are illuminating. It is found that commonly, as with the simpler model we have used before, $pH - \log [\alpha/(1 - \alpha)]$ is a nearly linear function of \bar{Z} for any given class of sites, where α and \bar{Z} have the same meaning as in (58). In such cases one may write, in analogy to (58)

$$pH = pk' + \log [\alpha/(1 - \alpha)] - 0.868w'\bar{Z} \quad (70)$$

Here pk' and w' are empirical parameters. It is found that the position of the sites, relative to the surface of the sphere, is crucial to the results obtained. For several different models examined, the value of w' was considerably less than w from (58)—taking the same values of b from (53)—if the charges were assumed to be on the surface. If they were assumed to be 2 or 3 Å inside the surface, w' became much greater than w . If w' is to be approximately equal to w , the charges must be taken as approximately 1 Å inside the surface of the sphere; and this assumption appears reasonably consistent with all the facts.

The value of pk' in (70) was also found to vary markedly for the different models chosen even if the same intrinsic pk values are assumed for the ionizing groups in the absence of electrostatic forces. This is not surprising; in some possible arrangements a given cationic group, for instance, may have as nearest neighbors other cationic groups; in other possible arrangements, anionic groups. Even if the total net charge on the molecule is the same in both cases, the group in question will have a lower pk' value in the former case than in the latter. Such effects have sometimes been interpreted in terms of hydrogen bonding (see for instance Loeb and Scheraga, 1956); they may equally well be due simply to electrostatic interactions of the type analyzed by Tanford. A final decision can be made only when knowledge of protein structure has become much more detailed than it is at present, and analysis of titration curves can then be carried out in terms of more precisely specified models.

REFERENCES

- Alfrey, T., Berg, P. W., and Morawetz, H. (1951). *J. Polymer Sci.* **7**, 543.
 Anfinsen, C. B., and Redfield, R. R. (1956). *Advances in Protein Chem.* **11**, 1.
 Bamford, C. H., Elliott, A., and Hanby, W. E. (1956). "Synthetic Polypeptides," Academic Press, New York.
 Benesch, R. E., and Benesch, R. (1955). *J. Am. Chem. Soc.* **77**, 5877.
 Brown, H. C., McDaniel, D. H., and Häfliger, O. (1955). Dissociation constants, in "Determination of Organic Structures by Physical Methods" (E. A. Braude and F. C. Nachod, eds.), p. 567, Academic Press, New York.
 Cannan, R. K., Kibrick, A. C., and Palmer, A. H. (1941). *Ann. N. Y. Acad. Sci.* **41**, 243.
 Cannan, R. K., Palmer, A. H., and Kibrick, A. C. (1942). *J. Biol. Chem.* **142**, 803.
 Carr, C. W. (1953). *Arch. Biochem. and Biophys.* **46**, 417.

- Cohn, E. J., and Edsall, J. T. (1943). "Proteins, Amino Acids and Peptides," Chapter 20, Reinhold Publishing Corp., New York.
- Cohn, E. J., Green, A. A., and Blanchard, M. H. (1937). *J. Am. Chem. Soc.* **59**, 509.
- Crammer, J. L., and Neuberger, A. (1943). *Biochem. J.* **37**, 302.
- Ebert, L. (1926). *Z. Physik. Chem.* **121**, 385.
- Fevold, H. L. (1950). *Advances in Protein Chem.* **6**, 188.
- Field, E. O., and O'Brien, J. R. P. (1955). *Biochem. J.* **60**, 656.
- Foster, J. F., and Aoki, K. (1957). *J. Phys. Chem.* **61**, 1369.
- Fraenkel-Conrat, H., and Mecham, D. K. (1949). *J. Biol. Chem.* **177**, 477, and earlier papers.
- French, D., and Edsall, J. T. (1945). *Advances in Protein Chem.* **2**, 277.
- Fromageot, C., and Schnek, G. (1950). *Biochim. et Biophys. Acta* **6**, 113.
- Fuoss, R. M., Katchalsky, A., and Lifson, S. (1951). *Proc. Natl. Acad. Sci. (U. S.)* **37**, 579.
- Grafius, M. A., and Neilands, J. B. (1955). *J. Am. Chem. Soc.* **77**, 3389.
- Harris, F. E., and Rice, S. A. (1955). *J. Phys. Chem.* **58**, 725.
- Harris, F. E., and Rice, S. A. (1956). *J. Chem. Phys.* **24**, 336.
- Henderson, L. J. (1928). "Blood: A Study in General Physiology," Yale University Press, New Haven, Connecticut.
- Hermans, J. J., and Overbeek, J. T. G. (1948). *Rec. trav. chim.* **67**, 761.
- Hill, T. L. (1955). *Arch. Biochem. and Biophys.* **57**, 229.
- Hughes, W. L. (1954). The proteins of blood plasma, in "The Proteins" (H. Neurath and K. Bailey, eds.), Vol. IIB, Chapter 21, Academic Press, New York.
- Katchalski, E. (1950). *Advances in Protein Chem.* **6**, 123.
- Katchalsky, A. (1954a). *Progr. in Biophys. and Biophys. Chem.* **4**, 1.
- Katchalsky, A. (1954b). *J. Polymer Sci.* **12**, 159.
- Katchalsky, A., and Lifson, S. (1953). *J. Polymer Sci.* **11**, 409.
- Katchalsky, A., and Miller, I. R. (1954). *J. Polymer Sci.* **13**, 57.
- Katchalsky, A., and Zwick, M. (1955). *J. Polymer Sci.* **16**, 221.
- Katchalsky, A., Shavit, N., and Eisenberg, H. (1954). *J. Polymer Sci.* **13**, 69.
- Klotz, I. M. (1953). Protein interactions, in "The Proteins" (H. Neurath and K. Bailey, eds.), Vol. IB, Chapter 8, Academic Press, New York.
- Laskowski, M., Jr., and Scheraga, H. A. (1954). *J. Am. Chem. Soc.* **76**, 6305.
- Lifson, S., and Katchalsky, A. (1954). *J. Polymer Sci.* **13**, 43.
- Linderstrom-Lang, K. (1924). *Compt. rend. trav. lab. Carlsberg (Sér. chim.)* **15**, No. 7.
- Loeb, G. I., and Scheraga, H. A. (1956). *J. Phys. Chem.* **60**, 1633.
- McMeekin, T. L. (1954). Milk proteins, in "The Proteins" (H. Neurath and K. Bailey, eds.), Vol. IIA, p. 389, Academic Press, New York.
- von Muralt, A. (1930). *J. Am. Chem. Soc.* **52**, 3518.
- Neuberger, A. (1936). *Biochem. J.* **30**, 2085.
- Overbeek, J. T. G. (1948). *Bull. soc. chim. Belges* **57**, 252.
- Pals, D. T. F., and Hermans, J. J. (1952). *Rec. trav. chim.* **71**, 458.
- Putnam, F. W. (1953). Chemical modification of proteins, in "The Proteins" (H. Neurath and K. Bailey, eds.), Vol. IB, p. 893, Academic Press, New York.
- Rice, S. A., and Harris, F. E. (1955). *J. Phys. Chem.* **58**, 733.
- Rice, S. A., and Harris, F. E. (1956). *J. Chem. Phys.* **24**, 326.
- Rykland, L. R., and Schmidt, C. L. A. (1944). *Arch. Biochem.* **5**, 89.
- Scatchard, G. (1949). *Ann. N. Y. Acad. Sci.* **51**, 660.
- Scatchard, G., Scheinberg, I. H., and Armstrong, S. H., Jr. (1950). *J. Am. Chem. Soc.* **72**, 535, 540.

- Sela, M., and Katchalski, E. (1956). *J. Am. Chem. Soc.* **78**, 3987.
- Sela, M., Katchalski, E., and Gahatia, M. (1956). *J. Am. Chem. Soc.* **78**, 746.
- Shugar, D. (1952). *Biochem. J.* **52**, 142.
- Simms, H. S. (1926). *J. Am. Chem. Soc.* **48**, 1231.
- Steinhardt, J., and Zaiser, E. M. (1951). *J. Biol. Chem.* **190**, 197.
- Steinhardt, J., and Zaiser, E. M. (1953). *J. Am. Chem. Soc.* **75**, 1599.
- Steinhardt, J., and Zaiser, E. M. (1955). *Advances in Protein Chem.* **10**, 151.
- Tanford, C. (1955). Hydrogen ion titration curves of proteins, in "Electrochemistry in Biology and Medicine" (T. Shedlovsky, ed.), p. 248, John Wiley & Sons, New York.
- Tanford, C. (1957a). *J. Am. Chem. Soc.* **79**, 3931.
- Tanford, C. (1957b). *J. Am. Chem. Soc.* **79**, 5340, 5348.
- Tanford, C., and Buzzell, J. G. (1956). *J. Phys. Chem.* **60**, 225.
- Tanford, C., and Epstein, J. (1954). *J. Am. Chem. Soc.* **76**, 2170.
- Tanford, C., and Hauenstein, J. D. (1956a). *J. Am. Chem. Soc.* **78**, 5287.
- Tanford, C., and Hauenstein, J. D. (1956b). *Biochim. et Biophys. Acta* **19**, 535.
- Tanford, C., and Kirkwood, J. G. (1957). *J. Am. Chem. Soc.* **79**, 5333.
- Tanford, C., and Roberts, G. L., Jr. (1952). *J. Am. Chem. Soc.* **74**, 2509.
- Tanford, C., and Wagner, M. L. (1954). *J. Am. Chem. Soc.* **76**, 3341.
- Tanford, C., Buzzell, J. G., Rands, D. G., and Swanson, S. A. (1955a). *J. Am. Chem. Soc.* **77**, 6421.
- Tanford, C., Hauenstein, J. D., and Rands, D. G. (1955b). *J. Am. Chem. Soc.* **77**, 6409.
- Tanford, C., Swanson, S. A., and Shore, W. S. (1955c). *J. Am. Chem. Soc.* **77**, 6414.
- Tristram, G. R. (1953). In "The Proteins" (H. Neurath and K. Bailey, eds.), Vol. IA, Chapter 3, Academic Press, New York.
- Warner, R. C. (1954). Egg proteins, in "The Proteins" (H. Neurath and K. Bailey, eds.), Vol. IIA, p. 435, Academic Press, New York.
- Weber, H. H. (1957). The Motility of Muscle and Cells. Dunham Lectures, Harvard University Press, Cambridge, Mass.
- Weber, H. H., and Portzehl, H. (1952). *Advances in Protein Chem.* **7**, 161.
- Weber, H. H., and Portzehl, H. (1954). *Progress in Biophys. and Biophys. Chem.* **4**, 60.
- Wyman, J. (1939). *J. Biol. Chem.* **127**, 1.
- Wyman, J. (1948). *Advances in Protein Chem.* **4**, 407.
- Yang, J. T., and Foster, J. F. (1954). *J. Am. Chem. Soc.* **76**, 1588.

Chapter 10

Carbon Dioxide and Carbonic Acid

Carbon Dioxide in Nature

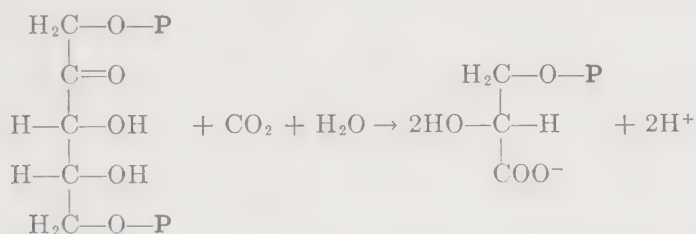
Carbon dioxide is ubiquitous and plays a central role both in the world at large and in the living organism. It is unique among gases in distributing itself almost equally, in concentration per unit volume, between air and water at ordinary temperatures. Hence it is undergoing perpetual exchange between the atmosphere and the waters of the world and between the living organism and its surroundings. It is unusual, though not unique, in being reversibly hydrated to form an acid, carbonic acid (H_2CO_3), in a process which is not instantaneous but requires a measurable time. This acid is dibasic, forming the bicarbonate (HCO_3^-) and carbonate (CO_3^{--}) ions as conjugate bases. The combination of the acidity and buffering power of H_2CO_3 with the volatility of CO_2 provides a mechanism of unrivaled efficiency for maintaining constancy of $p\text{H}$ in systems which are constantly being supplied, as living organisms are, with acidic products of metabolism.

Carbon dioxide is the great source of organic compounds in photosynthesis, the end product of respiration. An average man may produce a kilogram or more of metabolic carbon dioxide in the course of a day. To excrete such an amount, as Henderson has remarked, would present most formidable problems to the organism if it were not for the volatility of this substance. Such production of carbon dioxide by living organisms is proceeding on a vast scale throughout the world. It has been estimated by Goldschmidt and by Kalle (see Rankama and Sahama, 1950, p. 542) that some 30 to 35 mg of carbon dioxide is produced annually, by respiration and decay, for each square centimeter of the earth's surface. The production by combustion of coal and petroleum is a further significant contributing factor, amounting perhaps to 4% of that due to respiration, and is certainly increasing at the present time. The contribution due to forest and prairie fires has been estimated by Kalle as of comparable magnitude. There is also a gradual influx of carbon dioxide into the atmosphere from volcanic action and from underground springs. The annual influx from this source is very small—probably less than one two-

thousandth of that produced by respiration and decay—but the total, over the whole vast stretch of geological time, must have been prodigious.

On the other hand, the consumption of carbon dioxide by photosynthesis is probably as great, or somewhat greater, than the production due to respiration and decay. Nevertheless, it is probably not quite enough, at the present time, to balance the total production including the industrial production of carbon dioxide by man and the effects of forest fires. There is evidence that the carbon dioxide content of the atmosphere is therefore very slowly rising. There are, however, other processes which result in a very slow removal of CO_2 from active circulation. This is constantly proceeding by the formation of carbonates—especially calcium carbonate—in sedimentary rock; and the carbon of carbon dioxide is also being slowly removed by deposition in the form of coal and oil. These slow processes of removal approximately balance the production of “juvenile carbon dioxide” by volcanoes and springs.

The pathway by which carbon dioxide is incorporated into organic molecules in photosynthesis has been largely revealed by the brilliant work of Calvin and his associates (Calvin 1956; Bassham and Calvin 1956). The crucial reaction involves the combination of CO_2 with the ketopentose derivative, ribulose diphosphate, to give two molecules of phosphoglyceric acid. Denoting a phosphate ester group by **P**, we may write:



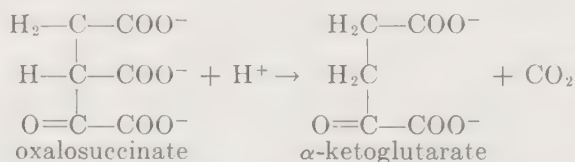
This reaction, which involves the formation of two new carboxyl groups per molecule of CO_2 incorporated, is only one link in a very elaborate chain of events, described in detail by Calvin. Most of the reactions in this cycle are found also in other tissues which are incapable of photosynthesis; only the reaction written above, and its immediate predecessor, in which ribulose monophosphate is phosphorylated to form the diphosphate, seem to be quite distinctive for photosynthetic tissues.

The biochemical *formation* of carbon dioxide is accomplished in numerous ways. In the Krebs tricarboxylic acid cycle, for instance, pyruvate reacts with the sulfhydryl group of coenzyme A (CoASH) to form acetyl coenzyme A (CoASCOCH_3), releasing CO_2 . The reaction may be regarded as an oxidative decarboxylation of pyruvate; the two electrons (ϵ^-) removed in the process may be accepted by one of the

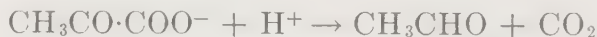
phosphopyridine nucleotides (DPN or TPN) or by other acceptors, including CO_2 itself, which may be reduced to formate ion.



In general the release of CO_2 in metabolic processes is accomplished by decarboxylations, as in several steps of the tricarboxylic acid cycle. These may be oxidative, as in the reaction shown above, or nonoxidative as in the conversion of oxalosuccinate to α -ketoglutarate.



Another notable nonoxidative decarboxylation is found in yeast fermentation, when pyruvate is converted to acetaldehyde by carboxylase, with diphosphothiamine as coenzyme:



the aldehyde being then reduced to ethanol. A comprehensive survey of the energy relationships in biochemical processes, treating these reactions among many others, has been given by Krebs and Kornberg (1957).

Respiration is indeed the great producer of carbon dioxide. In no organism, however, is this substance merely a breakdown product resulting from decomposition. It re-enters the active metabolic cycle in such fundamental processes as the formation of oxalacetate from pyruvate, which without postulating any particular mechanism may be written



This reaction may take place in either direction but only in the presence of enzymes. However, CO_2 also undergoes a reaction in the presence of amines to form carbamates. This requires no catalyst and is of definite biological importance:



Since all proteins contain free amino groups, carbamate formation must occur to some extent whenever CO_2 dissolves in a protein solution that is even mildly alkaline. It is responsible for a significant fraction of the carbon dioxide transported by the blood—a topic of which we shall have much more to say later.

Thus carbon dioxide, when it dissolves in a system containing other

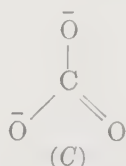
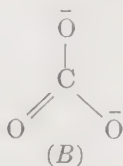
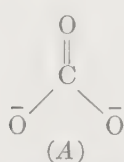
acids and bases, especially proteins, can be transformed into a variety of forms—dissolved CO_2 , H_2CO_3 , bicarbonate ion (HCO_3^-), carbonate ion (CO_3^{--}), carbamates, and perhaps other reaction products as well. All these forms exist in a mobile equilibrium with one another, and most of the reactions of interconversion proceed almost instantaneously. The formation of carbamates by reaction (2), although rapid, is not too rapid to measure, and its speed has been estimated by Faurholt (1925). Considerably slower than this is the reversible hydration of CO_2 to H_2CO_3 , the rate of which is measurable and has been measured with considerable accuracy in both directions. The relatively slow speed of this reaction is indeed a bottleneck in some important biochemical processes, notably in the uptake of carbon dioxide by the blood in the tissues of vertebrates, and its unloading from the blood in the lungs. Here, indeed, it is only the presence of the enzyme carbonic anhydrase in the red cells of the blood which permits these processes to go with the requisite speed. This enzyme is distinguished by containing zinc as an essential constituent; and the discovery of this fact by Keilin and Mann was the first specific evidence for a biochemical function of zinc, although zinc has since been shown to be an essential constituent of several other enzymes.

Structural Considerations

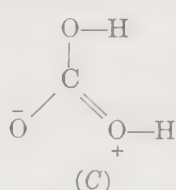
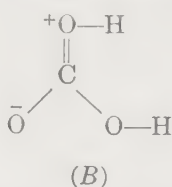
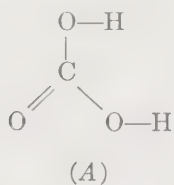
Carbon dioxide, like water, is a triatomic molecule. Unlike water, however, it is nonpolar, with a higher degree of symmetry, the three atoms being collinear and the two C—O bonds being equivalent. This is what might have been expected from the traditional formulation of the structure, in which each C—O bond was naturally supposed to be a double bond. Such double bonds, however, in the aldehydes and ketones are characterized by a C—O distance near 1.22 Å, whereas the C—O distances in CO_2 are found by spectroscopic measurements to be only 1.15 Å. This has been interpreted by Pauling (1940, p. 196) as due to resonance between the "classical" structure $\text{O}=\text{C}=\text{O}$ and other structures involving C—O triple bonds and charge displacement, such as $\text{O}\equiv\text{C}-\text{O}^-$ and $^-\text{O}-\text{C}\equiv\text{O}$. Either of the latter structures, if it existed as a separate entity, would of course be highly polar, but on account of symmetry they make equal and opposite contributions to the actual structure, which is nonpolar.

Carbonic acid (H_2CO_3) is quite a different structure. Its exact spatial configuration is not known, but it must be closely related to that of the carbonate ion, the dimensions of which are accurately known from studies on the crystal calcite (CaCO_3). The carbon atom and all three oxygens of the CO_3^{--} ion are coplanar, the oxygens forming an equilateral triangle around the carbon, each C—O distance being 1.31 Å. The structure may

be regarded as a resonance hybrid, with equivalent contributions from the three forms:



In H_2CO_3 and HCO_3^- there is of course no such perfect equivalence between the constituent structures of the resonating hybrid, and the three C—O bonds should no longer be all of the same length. Instead, of the three structures which presumably make the major contribution to the actual structure of H_2CO_3 , (A) is certainly of more importance than



(B) or (C), and the C—O bond will have more double-bond character and a shorter length in the oxygen atom which is not attached to hydrogen. Nevertheless it may be inferred from a knowledge of analogous compounds that all three C—O bonds possess a significant amount of double-bond character, and that this maintains the carbon and the three oxygens in a nearly coplanar configuration, and at angles not far from 120° to one another (Pauling, 1940, p. 208).

Thus the hydration of CO_2 to H_2CO_3 is a process requiring a rearrangement of valence bonds, the two C—O bonds of CO_2 , 180° apart and 1.15 Å long, being transformed to the three C—O bonds of H_2CO_3 , approximately 120° apart and not far from 1.3 Å long. We shall not attempt to comment here on the details of the electronic rearrangements that must be involved in the process, and indeed little is known of them. It is not surprising, however, that a process such as this should require an appreciable time, in contrast for example to such a process as the hydration of NH_3 to NH_4OH , in which the hydration process simply involves the formation of a hydrogen bond between the unshared electron pair in the ammonia molecule as acceptor, and one of the hydrogens of a water molecule as donor. Such a process involves no major alteration in the character of the other valence bonds in these molecules. Indeed in a complex of the type $\text{H}_3\text{N} : \cdots \text{H}-\text{O}-\text{H}$ it may become a matter of arbitrary definition depending on the $\text{N} : \cdots \text{H}$ distance, to specify whether a given ammonia molecule is hydrated or not. On the other hand, CO_2 and H_2CO_3 are two definite and distinct entities, and it is not surpris-

ing that the rates of transition between the two forms in water have been measured with considerable precision in recent years. The temperature coefficients of both the hydration and the dehydration reactions are considerable, indicating that an activation energy, of the order of 16 to 20 kcal per mole, is involved in the transition between CO_2 and H_2CO_3 .

The Fundamental Equilibria in Systems Containing Carbon Dioxide and Carbonic Acid

ABSORPTION COEFFICIENTS OF CARBON DIOXIDE IN WATER

According to Henry's law, the concentration of a gas in solution is directly proportional to the partial pressure of the gas in a vapor phase in equilibrium with that solution. The solubility coefficient, Q , in Table I

TABLE I
SOLUBILITY OF CARBON DIOXIDE IN WATER AND IN SODIUM CHLORIDE SOLUTIONS

Temperature (°C)	Q (in H_2O)	Q' (in 0.2 M NaCl)	A (in water)
0	0.0770	0.0726	1.725
10	0.0640	0.0610	1.247
20	0.0393	0.0373	0.942
25	0.0345	0.0328	0.841
30	0.0302	0.0286	0.750
35	0.0268	0.0255	0.674
38	0.0250	0.0238	0.636
40	0.0241	0.0229	0.615
50	0.0198	0.0190	0.519
100	0.244

Solubility in water is given in terms of the coefficient (Q) of Henry's Law:

$$Q = m_{\text{CO}_2}/p_{\text{CO}_2} \text{ (this is denoted as } S \text{ by Harned and Davis)}$$

$$m_{\text{CO}_2} = \text{total moles dissolved } \text{CO}_2 \text{ in all forms, per kilogram } \text{H}_2\text{O}$$

$$p_{\text{CO}_2} = \text{partial pressure of } \text{CO}_2 \text{ (in atmospheres)}$$

In sodium chloride solutions, the Henry's law coefficient (Q') is given by the ratio $Q' = Q/\alpha m$, where m is the molality of sodium chloride, and α is a "salting-out" factor. In later calculations in the text, p is expressed in millimeters of mercury, and the Henry's law coefficient becomes $q = Q/760$, or $q' = Q'/760$. The salting-out coefficient is given by the equation

$$\alpha = 0.1190 - 0.833 \times 10^{-3}t + 0.666 \times 10^{-5}t^2 \text{ (} t \text{ in } ^\circ\text{C)}$$

A gives the absorption coefficient of CO_2 in water, that is, the ratio (moles/liter in solution)/(moles/liter in gas phase) at the given temperature.

Data from H. S. Harned and R. Davis, *J. Am. Chem. Soc.* **65**, 2030 (1943); see also H. S. Harned and F. T. Bonner, *ibid.* **67**, 1026 (1945). The value of Q at 100° is from L. J. Henderson, "The Fitness of the Environment," The Macmillan Company, New York, 1913.

gives the concentration of the dissolved carbon dioxide in moles per kilogram of water at various temperatures when the partial pressure of carbon dioxide (p_{CO_2}) in equilibrium with the liquid is 1 atmosphere. Since Henry's law can be taken as holding with high precision, over a range of pressure below 1 atmosphere or thereabouts, the amount dissolved at any lower value of p_{CO_2} can be immediately calculated. As with virtually all gases dissolved in water, the solubility of CO_2 decreases with rising temperature. It also decreases with increasing salt concentration, the salting-out equation (Chapter 5, equation 51) holding with reasonable accuracy. The salting-out coefficient, denoted by α in Table I, is approximately 0.11 at ordinary temperatures. The figures of Harned and Davis for the solubility coefficient in 0.2 *M* sodium chloride are approximately 5% lower than the values in pure water at the same temperature.

TABLE II
ABSORPTION COEFFICIENTS (*A*) OF GASES AT 0°

Gas	<i>A</i>
Oxygen	0.049
Hydrogen	0.021
Nitrogen	0.024
Carbon monoxide	0.035
Carbon dioxide	1.725
Sulfur dioxide	79.8
Ammonia	1299

From L. J. Henderson, "The Fitness of the Environment," p. 137, The Macmillan Company, New York, 1913. The value for carbon dioxide is from Table I and differs slightly from that given by Henderson.

The ionic strength of blood plasma is near 0.16; so in a salt solution comparable to protein-free blood plasma the solubility coefficient should be about 4% less than in water. The proteins introduce additional complications, since they may bind some CO_2 in the form of carbamate, whereas on the other hand they may exert some additional salting-out effect. Probably, however, these effects in blood plasma are both small.

In pure water or in sodium chloride solutions, more than 99% of the dissolved carbon dioxide is actually in the form of CO_2 . The total concentration of H_2CO_3 , together with that of bicarbonate and carbonate ions, amounts to much less than 1% of the dissolved CO_2 at any temperature between 0° and 50°.

The significance of the absorption of carbon dioxide by water is for some purposes more directly perceived by considering the coefficient *A*, in the last column of Table I which gives the ratio of the volume concentrations of CO_2 in the liquid and in the vapor phase. It is very close to

unity, and is indeed equal to unity at a temperature slightly below 20°. Most other gases are far less evenly distributed between air and water; some examples are given in Table II. We have already remarked on the significance of this fact; in the words of Henderson: "Thus the waters can never wash carbonic acid out of the air, nor the air keep it from the waters. It is the one substance which thus, in considerable quantities relative to its total amount, everywhere accompanies water. In earth, air, fire, and water alike these two substances are always associated."

THE REVERSIBLE HYDRATION OF CARBON DIOXIDE AND THE IONIZATION OF CARBONIC ACID

Carbon dioxide is not an acid in the sense of Brönsted's definition (Chapter 8), since it contains no protons.¹ It becomes a dibasic acid on hydration to H_2CO_3 , which on ionization yields HCO_3^- and CO_3^{--} ions. In a system which contains bases which can accept protons from H_2CO_3 , large concentrations of these ions may be present. On acidifying the solution, however, all these different forms are converted into H_2CO_3 and CO_2 , and quantitatively removed from the system as gaseous CO_2 , which may be trapped and analyzed. We may therefore define a quantity, T , commonly called the "total carbon dioxide content" of the system, which is directly accessible to experimental measurement. For the time being we assume that the bases present in the system are not of the type capable of forming carbamino derivatives with CO_2 , according to reaction (2). Then the value of T becomes

$$T = (\text{CO}_2) + (\text{H}_2\text{CO}_3) + (\text{HCO}_3^-) + (\text{CO}_3^{--}) \quad (3)$$

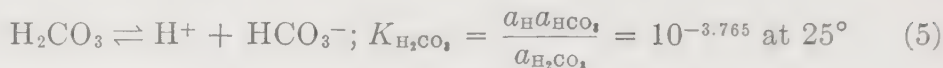
We now consider the interrelated system of equilibria between these different molecules and ions. There are four such equilibria to be considered, which we discuss in turn.

A. The Hydration Reaction



In defining K_h , the activity of water has been taken as unity. The numerical value of K_h given here is derived from those of the other equilibrium constants given below. It may also be derived from the ratio of the velocity constants for the hydration of CO_2 , and the dehydration of H_2CO_3 (Table VII). The methods of determining these are discussed later in some detail.

¹ Carbon dioxide is probably to be considered as an acid in the very general sense of the definition of G. N. Lewis. Its direct reaction with aliphatic amines to form carbamates (equation 2) may be somewhat analogous to the electrophilic attack of boron trichloride on ammonia to form the $\text{Cl}_3\text{B}-\text{NH}_3$ molecule (p. 410).

B. The Ionization of H_2CO_3 

Most of the determinations of $K_{H_2CO_3}$ have been made by indirect methods, using measured values of K_1 (defined below) from electromotive

TABLE III
VALUES OF pK_1 AND OF $pK_{H_2CO_3}$ FOR CARBONIC ACID, AND OF THE HYDRATION CONSTANT, K_h , AT VARIOUS TEMPERATURES IN WATER AND IN SODIUM CHLORIDE SOLUTION

Temperature (°C)	pK_1 (in water)	$pK_{1(NaCl)}$ in NaCl solution			$pK_{H_2CO_3}$ (in water)	K_h
		0.1 M	0.2 M	0.5 M		
0	6.578	6.349	6.288	6.206		
5	6.517	6.287	6.227	6.147	3.81	0.00200
10	6.464	6.234	6.174	6.095		
15	6.419	6.188	6.128	6.051	3.755	0.00212
20	6.381	6.148	6.089	6.013		
25	6.352	6.116	6.057	5.982	3.765	0.00258
30	6.327	6.089	6.030	5.958		
35	6.309	6.068	6.009	5.938	3.78	0.00296
38	6.302	6.059	6.000	5.929	3.80	0.00317
40	6.298	6.053	5.994	5.924		
45	6.290	6.034	5.984	5.814	3.795	0.00321

The value of K_1 (equation 6) is the limit, at zero ionic strength, of $(H^+)(HCO_3^-)/S$, where $S = (CO_2) + (H_2CO_3)$, as the concentrations of all these species approach zero. The value of $K_{1(NaCl)}$ is the same limit, but approached at a fixed constant ionic strength, determined by the molality of sodium chloride in the solvent. All values are on the molality scale. $K_{H_2CO_3}$ is the true first ionization constant of carbonic acid (equation 5).

Values of pK_1 in water and sodium chloride solutions from emf measurements of H. S. Harned *et al.* (see Table I). The original papers give considerably more data than those recorded here. Compare also T. Shedlovsky and D. A. MacInnes, *J. Am. Chem. Soc.* **57**, 1705 (1935); D. A. MacInnes and D. Belcher, *ibid.* **55**, 2630 (1933); **57**, 1683 (1935), for values obtained by conductivity measurements and with the glass electrode. Values of $pK_{H_2CO_3}$ from K. F. Wissbrun, D. M. French, and A. Patterson, Jr., *J. Phys. Chem.* **58**, 693 (1954). These pK values are probably reliable only to ± 0.012 , in contrast to those of pK_1 which may be reliable within ± 0.001 . Values of K_h calculated from K_1 and $K_{H_2CO_3}$ by equation (8). A. B. Hastings and J. Sendroy, Jr., *J. Biol. Chem.* **65**, 445 (1925) studied the pK' values of carbonic acid at 38° , using cells with liquid junction, as a function of ionic strength, ω , and found that their data could be described by the simple equations

$$\begin{aligned} pK_1' &= 6.33 - 0.5 \sqrt{\omega} \\ pK_2' &= 10.22 - 1.1 \sqrt{\omega} \end{aligned}$$

force or conductivity measurements, and of K_h determined by the ratio of the velocity constants for hydration and dehydration. Recently, however, Berg and Patterson (1953) determined $K_{\text{H}_2\text{CO}_3}$ at 25° directly from measurements of the conductance of carbonic acid solutions at high field strengths, and more accurate data by the same method have been obtained by Wissbrun *et al.* (1954) at a series of temperatures from 5° to 38° (see Table III). The equivalent conductance of a weak electrolyte is increased, at high field strengths, above its value in field of low intensity. This is due to two effects: (1) The ions move so fast, at high field strengths, that they partly escape from the retarding effects of the surrounding ion atmosphere (Chapter 7). This effect is observed in both strong and weak electrolytes. (2) In weak electrolytes, a very intense field actually shifts the equilibrium between ions and neutral molecules in favor of the ions. Roughly, we may think of the field as pulling the molecules apart, as rapidly as they are formed, and thereby increasing the degree of ionization. These two effects, known collectively as the Wien effect, can be used as a basis for determining ionization constants. The method involves the determination of the slope of the curve for conductance as a function of field strength. The fractional degree of ionization of the electrolyte at zero field strength, and hence the ionization constant, is a function of this slope. No attempt will be made to give the theory here, but the results for H_2CO_3 are given in Table III. It will be seen that carbonic acid is about ten times as strong an acid as acetic acid, nearly as strong as formic acid.

C. The "Over-All" First Ionization Constant



$$K_1 = \frac{a_{\text{H}}a_{\text{HCO}_3}}{a_{\text{CO}_2} + a_{\text{H}_2\text{CO}_3}} = a_{\text{H}}a_{\text{HCO}_3}/S = 10^{-6.352} \text{ at } 25^\circ \quad (6)$$

The equilibrium constant, K_1 , is what is determined by measuring the ion activity product, $a_{\text{H}}a_{\text{HCO}_3}$, from electromotive force or conductance measurements, in solutions equilibrated with a given partial pressure of carbon dioxide in the gas phase, so that the concentration of $\text{H}_2\text{CO}_3 + \text{CO}_2$ is fixed. The value of $a_{\text{CO}_2} + a_{\text{H}_2\text{CO}_3}$ is in the limit, at very low ionic strength, approached by the sum, S , of the concentrations of these substances:

$$S = (\text{CO}_2) + (\text{H}_2\text{CO}_3) = Qp_{\text{CO}_2} \quad (7)$$

Values of Q (or of Q' in salt solutions) are given in Table I. The constant K_1 is so important for many biochemical acid-base equilibria that values

for it at a number of temperatures are listed in Table III, together with limiting values for the ratio $(\text{H}^+)(\text{HCO}_3^-)/S$ in certain salt solutions.²

The very accurate data of Harned and his associates, a few of which are recorded in Table III, permit the calculation of the standard changes in free energy, heat content, entropy, and heat capacity, associated with the process described by pK_1 . These give the over-all changes associated with the reaction sequence $\text{CO}_2 + \text{H}_2\text{O} \rightleftharpoons \text{H}_2\text{CO}_3 \rightleftharpoons \text{H}^+ + \text{HCO}_3^-$.

Data are now available, however, which permit an evaluation for all these thermodynamic properties associated with the two separate steps involved: hydration and ionization. For the equilibrium constant and the standard free energy change we have, from (4), (5), and (6),

$$\frac{K_1}{K_{\text{H}_2\text{CO}_3}} = \frac{K_h}{K_h + 1} = \frac{10^{-6.352}}{10^{-3.765}} = 10^{-2.59} = 0.00257 \text{ at } 25^\circ \quad (8)$$

Hence $K_h = 0.00258$ at 25° ; $pK_h = 2.59$; and

$$\Delta F^\circ = 2.303RT pK_h = 3530 \text{ cal mole}^{-1}$$

Roughton (1941) has measured, by a rapid-flow method (see p. 581), the thermal changes associated with the reaction $\text{H}^+ + \text{HCO}_3^- \rightleftharpoons \text{H}_2\text{CO}_3$, which occurs when a bicarbonate solution is mixed with a strong acid, and also the further changes which occur as H_2CO_3 is dehydrated to CO_2 . The production of heat by the first reaction is virtually instantaneous; that of the second occurs much more slowly, although even this reaction is largely complete in a few hundredths of a second at 25° . Thus, these measurements give separately the heat changes in the combination of H^+ and HCO_3^- ions, and in the dehydration reaction.³ The sum of the two values checks satisfactorily with the total value of ΔH° , although neither figure considered separately is known as accurately as the sum of the two. Roughton determined the heat of reaction at several temperatures, from 0° to 38° , so that the heat capacity changes in the ionization and dehydration reactions can also be obtained from his data. The results are given in Table IV.

As is noted in the table, there are still some puzzling discrepancies between the temperature coefficient of $K_{\text{H}_2\text{CO}_3}$, as determined by Wissbrun

² We should call the reader's attention here to certain matters of notation. Roughton (1943-44) has used K_1 to denote what we have called $K_{\text{H}_2\text{CO}_3}$, and K_1' to denote what we have called K_1 . We have deviated from his notation, since we wish to reserve K_1' , as with other acids (Chapter 8), to denote the "apparent constant" $K_1' = a_{\text{H}}(\text{HCO}_3^-)/S$, involving a_{H} and the concentrations of acid and conjugate base; K_1 in our notation is defined by equation (6) in terms of the activities of the components.

³ Also, the time course of the heat evolution in the latter reaction permits a calculation of the velocity constant of the reaction. This point is considered later (p. 585) in connection with other methods for determining the same velocity constant.

TABLE IV
THERMODYNAMIC FUNCTIONS FOR HYDRATION OF CO₂ AND IONIZATION OF
CARBONIC ACID AT 25°

Equilibrium constant	pK	ΔF° cal mole ⁻¹	ΔH° cal mole ⁻¹	ΔS° cal deg ⁻¹ mole ⁻¹	ΔC_p° cal deg ⁻¹ mole ⁻¹
(H ₂ CO ₃)/(CO ₂) = K_h	2.59	3530	1130	- 8	-63
(H ⁺)(HCO ₃ ⁻)/(H ₂ CO ₃) = $K_{H_2CO_3}$	3.765	5170	1010	-14	-28
(H ⁺)(HCO ₃ ⁻)/ S = K_1	6.352	8666	2240	-21.6	-90
(H ⁺)(CO ₃ ⁻⁻)/(HCO ₃ ⁻) = K_2	10.329	14092	3603	-35.2	-65

Values of pK and ΔF° for $K_{H_2CO_3}$ from Wissbrun *et al.* (see Table III). Values for ΔH° and ΔC_p° for K_h and $K_{H_2CO_3}$ from F. J. W. Roughton, *J. Am. Chem. Soc.* **63**, 2930 (1941). Values for K_1 from Harned and Davis (see Table I); for K_2 from H. S. Harned and S. R. Scholes, *J. Am. Chem. Soc.* **63**, 1706 (1941). The sum of the ΔH° values for K_h and $K_{H_2CO_3}$ should be equal to ΔH° for K_1 ; and the same should be true for ΔC_p° . The figures show the extent of agreement between the calorimetric values of Roughton and the entirely independent estimates derived by Harned and Davis from the temperature variation of K_1 determined by electromotive force measurements.

The values for ΔH° and ΔC_p° derived from temperature variation of the $pK_{H_2CO_3}$ values of Wissbrun *et al.* (Table III) do not agree well with Roughton's calorimetric data. Thus Wissbrun *et al.* give $\Delta H^\circ = 2650$ for the hydration of CO₂ at 27°; Roughton gives 1040. The calculated values of ΔH° for $pK_{H_2CO_3}$, based on these different data, therefore differ also, but in the opposite direction. Tentatively we assume the calorimetric values to be the more reliable, but there is a puzzling discrepancy here which needs further work to resolve it.

et al. (1954), and the values to be expected from Roughton's thermal data; but the main picture seems clear.

D. The Second Ionization Constant

$$HCO_3^- \rightleftharpoons H^+ + CO_3^{--}; K_2 = \frac{a_H a_{CO_3}}{a_{HCO_3}} = 10^{-10.329} \text{ at } 25^\circ \quad (9)$$

Thermodynamic functions for this step of ionization are also given in Table IV.

Carbon Dioxide Dissociation Curves

The excretion of carbon dioxide by the vertebrates involves its passage from the tissues, where it is produced at a relatively high partial pressure, into the blood, and its subsequent unloading from the blood into the alveolar spaces of the lungs at a lower partial pressure. From the lungs, of course, the respiratory movements expel it into the atmosphere. If the process is to be efficient, a relatively small change in p_{CO_2} between tissues and lungs must result in a relatively large change in T , the total CO₂, as defined in equation (3). Any given portion of blood requires a

period of the order of 1 to 5 seconds to pass through a capillary in the lungs or tissues. In our present calculations we assume that equilibrium is attained in this time. This assumption corresponds very closely to the facts, because the catalytic activity of the enzyme carbonic anhydrase in the red cells speeds up the hydration of CO_2 in the tissue capillaries and the dehydration of H_2CO_3 in the lung capillaries to such an extent that all the reactions involved are essentially complete in this very short time.

Equilibrium studies on blood permit the determination of T as a function of p_{CO_2} , and the form of the experimental curves—generally called carbon dioxide dissociation curves—is similar to that of curve *B* in Fig. 1. On the other hand, a similar curve determined on blood plasma alone has a very different appearance, similar to curve *A* in Fig. 1. It is

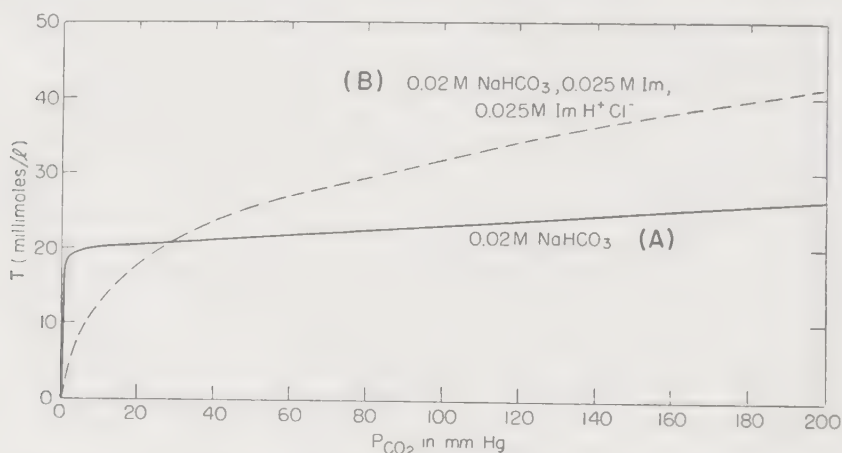


FIG. 1. Carbon dioxide dissociation curves for a sodium bicarbonate solution with and without an added imidazole buffer. Data from Tables V and VI.

obvious that whole blood is far more efficient than plasma as a vehicle for taking up and unloading CO_2 as the partial pressure varies, the relative efficiency of the two media being proportional to the slope of the dissociation curve in the physiologically important range, which is near $p_{\text{CO}_2} = 50$ mm Hg. This difference is not due to the fact that the system of plasma plus red cells is a two-phase system. If we prepare a single solution containing plasma proteins, hemoglobin, and inorganic ions, in the same total concentrations found in the same volume of whole blood, this solution has a carbon dioxide dissociation curve very like that of blood. Indeed a relatively simple system, containing no proteins or other macromolecules, can be prepared, which closely parallels the behavior of blood at a fixed oxygen pressure and varying p_{CO_2} .⁴ We shall proceed to examine the

⁴ In the present discussion we leave out of consideration the changes in the acid strength of hemoglobin which occur when it is oxygenated. These changes, which are of the greatest biological importance, are considered in Volume II.

requirement for the existence of such a system. First, however, we shall consider a still simpler system which behaves like blood plasma, rather than whole blood. In this discussion we neglect carbamate formation, which will be considered later.

In any system of this sort, two fundamental equations must be obeyed: first, equation (3) for T , the total CO_2 of the system, and second, the condition of electrical neutrality.

In terms of the ionization constants of carbonic acid, and of the partial pressure, p , of carbon dioxide in equilibrium with the vapor phase (see equations 6 to 9), T may be written

$$T = S + (\text{HCO}_3^-) + (\text{CO}_3^{--})$$

$$= q'p \left(1 + \frac{K_1'}{(\text{H}^+)} + \frac{K_1'K_2'}{(\text{H}^+)^2} \right) = 10^{-4.5}p \left(1 + \frac{10^{-6.1}}{(\text{H}^+)} + \frac{10^{-15.9}}{(\text{H}^+)^2} \right) \quad (10)$$

The numerical values chosen here require some comment. The partial pressure, p , is in millimeters of mercury. The values of S , (HCO_3^-) , and (CO_3^{--}) are expressed as concentrations, not as activities, since T is measured experimentally as a total concentration. Since we are considering a system analogous to mammalian blood, we choose values of q' , K_1' , and K_2' which are appropriate for the temperature (38°) and ionic strength ($\omega \cong 0.16$) of blood. The value of

$$q' = Q/760 = 0.0238/760 = 10^{-4.5}$$

for 0.2 M sodium chloride at 38° (Table I) should be a close approximation to the value for plasma. The values of pK_1 and pK_2 at zero ionic strength and 38° are 6.305 and 10.230, respectively. At ionic strength 0.16, the apparent pK' values, defined in terms of concentrations of acid and conjugate base, are lower than the limiting pK values expressed in terms of the activities of these species. The activity of each species may be expressed as the product of its concentration times its activity coefficient, denoted by γ with the appropriate subscript:

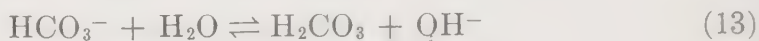
$$p\text{H} = pK_1' + \log \frac{(\text{HCO}_3^-)}{S} = pK_1 + \log \frac{(\text{HCO}_3^-)}{S} + \log \frac{\gamma_{\text{HCO}_3^-}}{\gamma_{\text{CO}_2}} \quad (11)$$

$$p\text{H} = pK_2' + \log \frac{(\text{CO}_3^{--})}{(\text{HCO}_3^-)} = pK_2 + \log \frac{(\text{CO}_3^{--})}{(\text{HCO}_3^-)} + \log \frac{\gamma_{\text{CO}_3^{--}}}{\gamma_{\text{HCO}_3^-}} \quad (12)$$

The appropriate values for pK_1' and pK_2' for $\omega = 0.16$ may be estimated from equations (71) and (73) of Chapter 8. Since the calculations to be given here are primarily to illustrate the principles involved, we specify the appropriate values of pK' only within 0.1, taking $pK_1' = 6.1$ and $pK_2' = 9.8$. For the ionization constant of water, which will be needed in

the calculations, we find from Table II of Chapter 8, $pK_w = 13.6$ at 38° . Actually the form of the curves to be discussed here will be much the same, regardless of minor variations in the values assigned to these constants.

As a simple example of a system somewhat analogous to blood plasma we may consider, for instance, a system containing C moles of sodium bicarbonate in a liter of water, in equilibrium with a variable pressure, p , of CO_2 in the vapor phase. Qualitatively it is clear that, as p is diminished, CO_2 and H_2CO_3 are removed from the solution. In the process, H_2CO_3 and carbonate ions are formed from bicarbonate, by interaction with water. The results may be described by the equations



The H_2CO_3 is dehydrated to form CO_2 and passes off into the vapor phase; these equations thus represent the loss of half the CO_2 of the bicarbonate as H_2CO_3 , the other half being converted to carbonate ion.

As the pressure of CO_2 falls to still lower levels, the process occurring may be formulated as an interaction of the carbonate ion with water according to the reverse of equation (14):



The HCO_3^- ions thus formed interact according to (13), evolving more CO_2 and OH^- ion. Finally, at zero pressure of CO_2 , all the CO_2 is removed from the system, and the sodium bicarbonate is completely converted to sodium hydroxide.

We may now consider the course of these reactions quantitatively, first formulating the equation of electrical neutrality. The cations present are sodium and hydrogen ion, the concentration of the former being fixed at the value C . The anions present are HCO_3^- , CO_3^{--} (with two negative charges), and OH^- . Hence:

$$\begin{aligned} (\text{Na}^+) + (\text{H}^+) &= C + (\text{H}^+) = (\text{HCO}_3^-) + 2(\text{CO}_3^{--}) + (\text{OH}^-) \quad (15) \\ &= (\text{HCO}_3^-) + 2(\text{CO}_3^{--}) + \frac{K_w}{(\text{H}^+)} \end{aligned}$$

or, employing (10) and (11) for the values of (CO_3^{--}) and (HCO_3^-) :

$$\begin{aligned} C + (\text{H}^+) &= \frac{K_1'q'p}{(\text{H}^+)} + \frac{2K_1'K_2'q'p}{(\text{H}^+)^2} + \frac{K_w}{(\text{H}^+)} \\ &= \left[\frac{10^{-10.6}}{(\text{H}^+)} + \frac{10^{-20.1}}{(\text{H}^+)^2} \right] p + \frac{10^{-13.6}}{(\text{H}^+)} \quad (16) \end{aligned}$$

This gives for p in terms of (H^+) and the K' values:

$$p = \frac{C + (\text{H}^+) - K_w/(\text{H}^+)}{K_1'q/(\text{H}^+) + 2K_1'K_2'q/(\text{H}^+)^2} = \frac{C + (\text{H}^+) - 10^{-13.6}/(\text{H}^+)}{10^{-10.6}/(\text{H}^+) + 10^{-20.1}/(\text{H}^+)^2} \quad (17)$$

The numerical values of K_1' , K_2' , and q have been inserted on the right-hand side of this equation. We may now insert values of (H^+) into (17), and solve for the corresponding values of p . From this value of p , and the associated value of (H^+) , T may be obtained from equation (10). We may consider some numerical results for a system containing 0.02 M sodium bicarbonate in water. This is not far from the concentration of bicarbonate ion in blood. Setting $C = 0.02$ in (17), we may proceed to evaluate p as a function of (H^+) . The actual use of the equation may be greatly simplified in many cases by dropping out negligible terms. We are concerned only with $p\text{H}$ values greater than 5. Hence $(\text{H}^+) < 10^{-5}$ in all cases, and this is negligible in comparison with C in the numerator. Also for $(\text{H}^+) > 10^{-9.6}$ the term $K_w/(\text{H}^+)$ in the numerator is negligible for most purposes. If $(\text{H}^+) \geq 10^{-8}$, the second term in the denominator is negligible in comparison with the first. If $(\text{H}^+) = 10^{-9.6}$ or less, the second term in the denominator is the larger, although the first is by no means negligible. If $(\text{H}^+) = K_w/C$ —a condition attained in this case at $p\text{H}$ 11.9— p becomes practically zero; that is, at this point approximately all the CO_2 has been removed, and the sodium bicarbonate completely converted into sodium hydroxide. Numerical relations between $p\text{H}$, p_{CO_2} , and total CO_2 are listed in Table V and plotted as curve A of Fig. 1.

TABLE V

TOTAL CO_2 (T) AS A FUNCTION OF p_{CO_2} AND $p\text{H}$ IN A SYSTEM CONTAINING 0.02 M SODIUM BICARBONATE AT 38°

$p\text{H}$	p_{CO_2} (mm Hg)	T (mM/l)
6.1	639	39.8
6.6	200	26.3
7.1	63	22.0
7.4	31.6	20.9
7.6	20.0	20.6
8.1	6.1	19.8
8.6	1.77	18.9
9.1	0.45	17.1
9.6	8.8×10^{-2}	14.3
10.6	1.40×10^{-3}	10.2
11.6	7.9×10^{-6}	5.1
11.9	0	0

The most striking feature of the behavior of this system is the nearly constant value of T which is maintained over a very wide range of partial pressures of CO_2 . Between $p = 63$ and $p = 0.45$ mm Hg, T varies only from 22 to 17.1 mM per liter; and at $p\text{H } 10.6$, when p has fallen to 0.0014, T is still equivalent to half the molar concentration of sodium bicarbonate originally added. Indeed it has been frequently stated that only half the total CO_2 originally present can be pumped off from a sodium bicarbonate solution, or from blood plasma, by exposing it to a vacuum. If the experimental conditions in the vacuum correspond to $p_{\text{CO}_2} \cong 10^{-3}$ mm Hg, this statement would fit the data of Table V. In principle, however, if p_{CO_2} can actually be reduced to zero, T should become zero also. Actually it would be very difficult experimentally to reduce p_{CO_2} so nearly to zero that all, or nearly all, the total CO_2 could be thus removed.

The carbon dioxide dissociation curves of blood, as we have already remarked, are very different. The total CO_2 can all be pumped off without difficulty by exposure to a vacuum, and the slope of the curve for T as a function of p_{CO_2} , when p_{CO_2} is near 50 mm Hg, is much steeper than for the sodium bicarbonate solution. The difference is due primarily to the presence of hemoglobin in blood, in high concentration, and to the unusually great buffer capacity of hemoglobin, in the physiological $p\text{H}$ range, as compared with most other proteins. This is readily explained by the fact that hemoglobin (molecular weight 66,700) contains thirty-six histidine residues per molecule (Chapter 3, Table II—value for horse hemoglobin), and that the imidazole groups of the histidine residues have pK values not far from 7 (Chapter 8, Table VI). Also, horse hemoglobin contains six terminal α -amino groups of valine residues, at the ends of peptide chains (Sanger, 1952), and these should have pK_A' values not far from 7.8. Some twenty of these groups were found by Cohn *et al.* (1937) to ionize as if they could all be characterized by a pK' value near 7.4. The molar concentration of hemoglobin, per liter of whole blood, is near 0.002, so that there are $20 \times 0.002 = 0.040$ mole of imidazole or amino groups per liter with a pK' value near 7.4. In addition the plasma proteins contribute approximately 0.010 mole of imidazole and α -amino groups per liter, with pK' values in this range. A simple imidazole derivative, 4-methylimidazole, has a pK_A' value close to 7.4 at 38° . As a simple model system for the carbon dioxide dissociation curve of blood, then, we select a solution containing sodium bicarbonate (0.020 M) and a mixture of 4-methylimidazole (Im) and its hydrochloride (HIm^+Cl^-). The total concentration $[\text{Im}] + [\text{HIm}^+] = C_1$ is taken as 0.050, and at the start of the experiment we add Im and HIm^+Cl^- in equimolar proportions. Hence $(\text{Na}^+) = 0.020$, and $(\text{Cl}^-) = 0.025 = C_1/2$. The pK_A' value of the ImH^+ ion is the negative logarithm of the constant defined by the

equations

$$\frac{(\text{H}^+)(\text{Im})}{(\text{HIm}^+)} = K_{\text{Im}}' = 10^{-7.4} \quad (18)$$

$$(\text{HIm}^+) = \frac{(\text{H}^+) \cdot C_{\text{I}}}{K_{\text{Im}}' + (\text{H}^+)} \quad (19)$$

The equation for electrical neutrality is

$$[\text{Na}^+] + (\text{H}^+) + [\text{HIm}^+] = [\text{HCO}_3^-] + 2[\text{CO}_3^{--}] + \frac{K_w}{(\text{H}^+)} + [\text{Cl}^-] \quad (20)$$

With $(\text{Na}^+) = 0.020$, $(\text{Cl}^-) = 0.025$, and from equations (11) and (19), this becomes

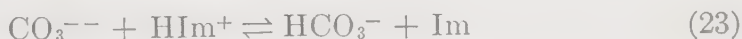
$$0.020 + (\text{H}^+) + \frac{0.050(\text{H}^+)}{K_{\text{Im}}' + (\text{H}^+)} = \frac{K_1'q'p}{(\text{H}^+)} + \frac{2K_1'K_2'q'p}{(\text{H}^+)^2} + \frac{K_w}{(\text{H}^+)} + 0.025 \quad (21)$$

and the partial pressure of CO_2 , p , is given as a function of (H^+) by

$$p = \frac{\frac{0.050(\text{H}^+)}{K_{\text{Im}}' + (\text{H}^+)} + (\text{H}^+) - \frac{K_w}{(\text{H}^+)} - 0.005}{\frac{K_1'q'}{(\text{H}^+)} + \frac{2K_1'K_2'q'}{(\text{H}^+)^2}} \quad (22)$$

When p is evaluated as a function of (H^+) , T may be calculated, as before, by equation (10).

This system behaves very differently from the sodium bicarbonate alone. If any CO_3^{--} ion is formed by reaction (12), as CO_2 is pumped off, it is immediately reconverted to HCO_3^- ion by HIm^+ ions present. The reaction may be written



The HCO_3^- ions thus formed then react (see equations 12 and 13) to form more CO_3^{--} ions and H_2CO_3 molecules. The latter pass off as CO_2 ; the former again interact with HIm^+ ions (equation 23). If the amount of HIm^+ present is more than sufficient to neutralize all the CO_3^{--} which can be formed—and this is true in the present case—then the reactions proceed until all the CO_2 in the system is driven off. The $p\text{H}$ of the system when all the CO_2 has been removed may readily be calculated from equation (22) by setting $p = 0$ in that equation. The condition for $p = 0$ is that the numerator of the expression on the right of (22) shall vanish. It is found that at all attainable values of p the terms (H^+) and $K_w/(\text{H}^+)$ are negligible in comparison with the other two terms in the

numerator. The approximate condition for $p = 0$ is then

$$\frac{0.050(\text{H}^+)}{K_{\text{Im}'} + (\text{H}^+)} = \frac{0.050(\text{H}^+)}{10^{-7.4} + (\text{H}^+)} = 0.005 \quad (24)$$

and this condition is attained at a value of (H^+) equal to 4.45×10^{-9} ($p\text{H} = 8.35$). This value is in striking contrast to the $p\text{H}$ value of 11.9 found for the sodium bicarbonate solution when all the CO_2 has been evacuated. The carbon dioxide dissociation curve for this system is given in Table VI and plotted as curve *B* of Fig. 1.

TABLE VI
CARBON DIOXIDE DISSOCIATION CURVE OF THE SYSTEM 0.02 *M*
SODIUM BICARBONATE—0.025 *M* Im—0.025 *M* ImH⁺Cl[−], AT 38°
(Im denotes 4-methylimidazole)

<i>pH</i>	<i>p</i> CO ₂ (mm Hg)	<i>T</i> (mM/l)
6.0	1725	98
6.6	381	50.2
7.0	123	34.6
7.2	64	27.5
7.4	31.7	21.7
7.6	14.1	14.7
7.8	5.7	9.3
8.0	2.0	5.14
8.2	0.45	1.86
8.35	0	0

Data calculated from equations (22) and (10), with $pK(\text{ImH}^+) = 7.4$. For acid-base equilibria in solutions of 4-methylimidazole, see Y. Nozaki, F. R. N. Gurd, R. F. Chen, and J. T. Edsall, *J. Am. Chem. Soc.* **79**, 2123 (1957).

Relations such as that shown in Table VI were recognized as long ago as 1907 by L. J. Henderson, who studied a system composed of carbon dioxide, bicarbonate ion, and primary and secondary phosphates. The combination of the carbonic acid–bicarbonate system with any other buffer mixture including a weak acid with a pK_A value near 7 would obviously be suitable for constructing such a system as we have described. We have chosen a system involving imidazole groups in our illustrative discussion, because of the physiological importance of these groups in hemoglobin and other proteins.

It is apparent from Table VI that systems of this sort are far more efficient in taking up and giving off carbon dioxide than the simple bicarbonate system. Near $p\text{H}$ 7.4, a change of 0.2 in $p\text{H}$ results in a change of approximately 6 mM per liter in the total carbon dioxide, for the system described in Table VI. Actually whole blood—or a hemoglobin solu-

tion—is much more efficient than this system could be in transporting carbon dioxide from the tissues to the lungs, and unloading it in the lungs. The oxygenation of hemoglobin increases the strength of certain acid groups in the molecule—commonly called the heme-linked acid groups. The increase of acidity which thus occurs in a hemoglobin solution when it is oxygenated in the lungs helps to drive off more carbon dioxide, and also serves to counterbalance the decrease in acidity due to the loss of

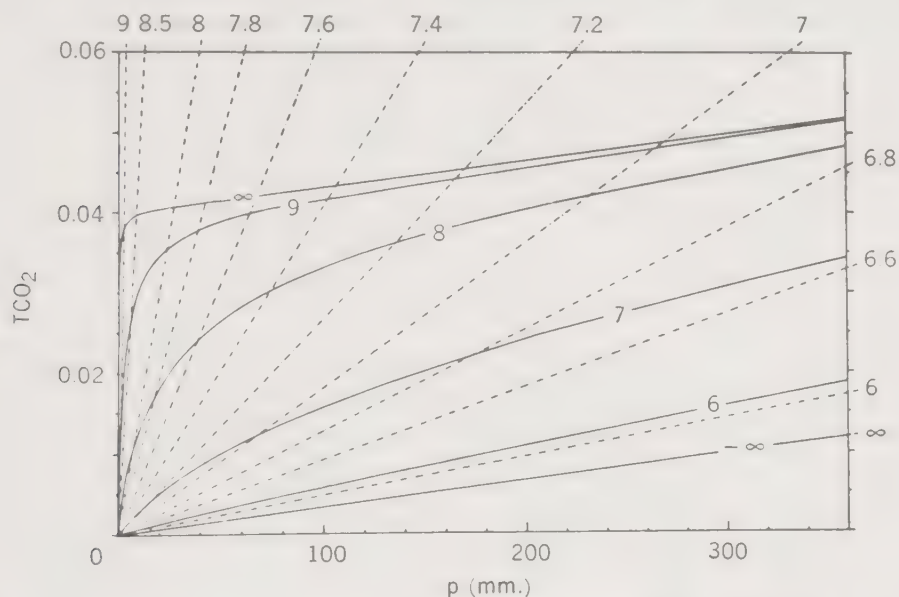


FIG. 2. Carbon dioxide dissociation curves for a system containing Na^+ ion, $0.04\text{ }M$, plus a buffer system of an acid HA and its conjugate base A , at a total concentration $(\text{HA}) + (\text{A}) = 0.06\text{ }M$. Solid contour lines show CO_2 dissociation curves for various assumed values of pK_A for the acid HA . The bottom line (denoted by $-\infty$) is for an infinitely strong acid HA and corresponds to the solution of pure CO_2 in water. The top solid line (denoted by ∞) is for an infinitely weak acid HA , i.e., for a solution consisting simply of $0.04\text{ }M$ sodium bicarbonate in water. The dotted contour lines are lines of constant $p\text{H}$; i.e., of constant $(\text{CO}_2)/(\text{HCO}_3^-)$ ratio.

carbon dioxide. The reverse relations occur in the tissues, and the $p\text{H}$ of the red cells may thus be maintained practically constant throughout a respiratory cycle. The detailed consideration of this beautifully organized system belongs in another chapter, but any consideration of carbon dioxide transport would be incomplete without referring to it.

In general, systems such as that analyzed in Table VI can be constructed by making up solutions containing x moles of sodium hydroxide, and C_A moles of an acid HA , per liter, and equilibrating them with carbon dioxide at various partial pressures. The total concentration $C_A = (\text{HA}) + (\text{A}^-)$ remains fixed. We denote K_A as the dissociation constant of HA . Then if HA is an uncharged acid, with an anion as

conjugate base, the equation of electrical neutrality becomes, noting that $x = (\text{Na}^+)$:

$$\begin{aligned} x + (\text{H}^+) &= (\text{A}^-) + (\text{HCO}_3^-) + 2(\text{CO}_3^{--}) + K_w/(\text{H}^+) \\ &= C_A K_A / [(\text{H}^+) + K_A] + (\text{HCO}_3^-) + 2(\text{CO}_3^{--}) + K_w/(\text{H}^+) \quad (21.1) \end{aligned}$$

If the acid is a cationic acid, such as ImH^+ , then it must be added with an anion, such as Cl^- , to balance its net charge. In that case $(\text{Cl}^-) = C_A$, and the reader can readily convince himself that equation (21.1) still holds (compare equation 21). Given this equation, the calculation of the

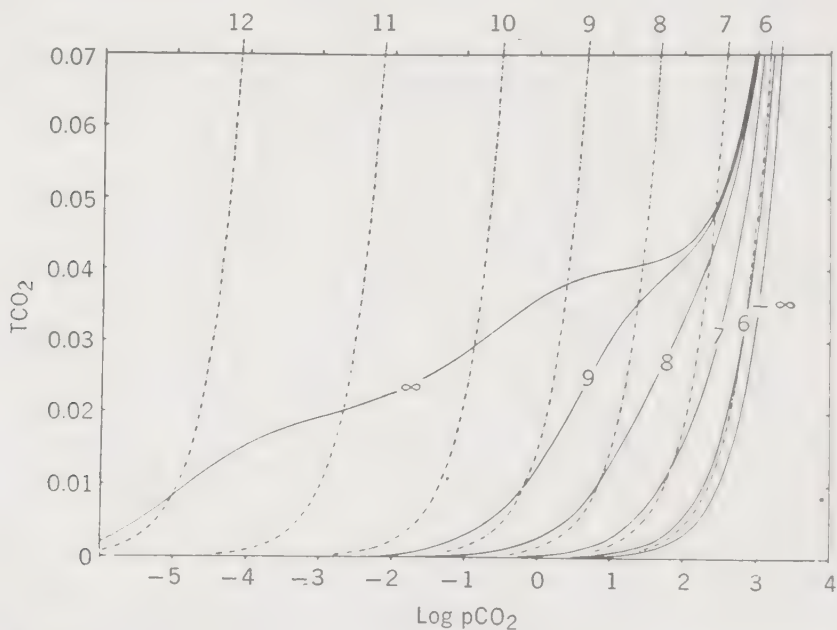


FIG. 3. The same data shown in Fig. 2, but with the abscissa taken as $\log p_{\text{CO}_2}$, in order to cover a wider range of partial pressures. As in Fig. 2, numerals attached to solid contour lines denote the pK_A value of the buffer acid HA. Dotted lines are lines of constant pH, with pH values attached.

carbon dioxide dissociation curves proceeds just as before. If C_A is appreciably greater than x , such systems function similarly to that shown in Table VI.

Figures 2 and 3 illustrate a set of such curves, for different values of pK_A , where we have taken $x = 0.04$ and $C_A = 0.06$. If the acid, HA, is completely dissociated to begin with—that is, an infinitely strong acid, corresponding to the curve marked $-\infty$ in Figs. 2 and 3—then the system reacts essentially like a solution of pure $\text{CO}_2 + \text{H}_2\text{CO}_3$ in water. If the acid is infinitely weak, as illustrated by the curve marked ∞ in the figures, then the system is equivalent to a sodium bicarbonate solution, such as that already discussed. For intermediate values of pK_A of the buffer acid a family of curves is obtained, some of which are shown in

the figures. It will be noted from Fig. 2 that the maximum slope of the curve, when p_{CO_2} is near 50 mm, is obtained when pK_A lies somewhere between 7 and 8. Such a buffer therefore would give maximum efficiency in the uptake and release of carbon dioxide as the partial pressure is varied in this range. For a system operating over a lower range of partial pressures of carbon dioxide, a system with a somewhat higher pK_A value for the buffer acid would be more efficient. Important discussions of carbon dioxide dissociation curves are given by Henderson (1928) and by Peters and van Slyke (1932).

The systems we have just considered bind no carbon dioxide in the form of carbamate. Since blood proteins—indeed proteins in general—and other physiologically important compounds do form carbamates, we now turn to the relations governing the formation and dissociation of these important compounds.

The Formation of Carbamates

The conversion of carbon dioxide to carbamates in solutions of compounds containing amino groups is a reaction of chemical interest and biological importance. No attempt will be made here to discuss it in great detail; for further information the reader should consult Faurholt (1925), Ferguson and Roughton (1934), Roughton (1935, 1943–44), Stadie and O'Brien (1935–36, 1937), Wyman (1948, pp. 485–496), and other references cited by them.

The general characteristics of the reactions involved may be briefly stated (see Roughton, 1943–44).

1. Carbon dioxide reacts with ammonia and the primary and secondary amino groups of aliphatic amines, amino acids, and proteins, according to equation (2):



(In primary amino groups, $\text{R}' = \text{H}$.) The reaction is rapid, much more rapid than the hydration of CO_2 to H_2CO_3 , and requires no catalyst. Direct reaction occurs only with uncharged amino groups, not with their conjugate acids, $\text{RR}'\text{NH}_2^+$. In simple amino acids, therefore, the dipolar ion is unreactive until a proton has been displaced from it; only the anion reacts.

2. The reactive species appears to be CO_2 . There is no evidence that H_2CO_3 , HCO_3^- , or CO_3^{--} reacts directly with amino groups.

3. Reaction (2) as written is strongly exothermic; rise of temperature therefore shifts the equilibrium to the left and causes the carbamino compound to dissociate into the amine and CO_2 .

4. The acids, $\text{RR}'\text{N}\cdot\text{COOH}$, are very unstable and decompose rapidly to the amine and CO_2 . Therefore carbamino compounds are decomposed

by addition of acid, but are stable in alkaline solution. Roughton (1941) studied the simplest member of the series, carbamic acid ($\text{H}_2\text{N}\cdot\text{COOH}$), and estimated its half-life to be of the order of 0.01 second at 0° .

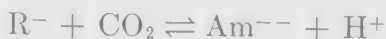
5. Roughton also estimated the pK value of the carboxyl group in carbamic acid to be near 5.8, about one pH unit above that of acetic acid. Thus formation of stable carbamates in significant amounts is unlikely to occur below pH 7. The weakening effect on acidity of the NH_2 group immediately adjoining the carboxyl is not surprising, since it makes possible the occurrence of resonance with forms such as $^+\text{H}_2\text{N} = \text{C}(\text{O}^-)\text{OH}$, in which the negative charge on one oxygen should certainly tend to bind the proton on the carboxyl group more tightly.⁵

6. The calcium and barium salts of carbamic acid are quite soluble in water, unlike calcium or barium carbonate. Thus in a solution containing both carbonates and carbamates, the former may be precipitated by addition of a calcium or barium salt, leaving the latter in solution. Analytical methods for estimating carbamino compounds have been based on this fact; see especially the work of Faurholt (1925), of Ferguson and Roughton (1934), and of Ferguson (1936).

In the light of this knowledge, consider the nature of the equilibria in a solution containing dissolved carbon dioxide and a simple amino acid such as glycine with one amino and one carboxyl group. Since the isoelectric point of such an amino acid is near pH 6, and since carbamate formation at equilibrium is relatively unimportant below pH 7, we can disregard the cationic form of the amino acid, which is present in significant amount only in acid solutions below pH 4.5. The only form of the amino acid which reacts directly with CO_2 is the anion, $\text{H}_2\text{N}\cdot\text{R}\cdot\text{COO}^-$, which we denote as R^- :



or



The equilibrium constant of this reaction is

$$K_{\text{Am}'} = \frac{(\text{Am}^{--})(\text{H}^+)}{(\text{R}^-)(\text{CO}_2)} = \frac{(\text{Am}^{--})(\text{H}^+)}{(\text{R}^-)q'p} \quad (26)$$

The equilibrium between the dipolar ion and the anion of the amino acid must also be considered (see Chapter 8, Table III, and the discussion in Chapter 9):

⁵ By contrast, the carboxyl group in $\text{H}_2\text{N}\cdot\text{CH}_2\cdot\text{COOH}$ (see the pK_a value in Chapter 9, equation 24) has a pK value of 4.4, and is thus more acid than acetic acid. In this structure, of course, such resonance as that found in carbamic acid cannot occur, since the intervening CH_2 group serves as an effective insulator.

$$K_2' = (H^+)(R^-)/(R^\pm) = 10^{-9.8} \text{ at } 25^\circ \quad (27)$$

The total concentration, C_A , of the amino acid in all forms (disregarding the cation for reasons already given) may be written as the sum of the concentrations of the dipolar ion, the anion, and the carbamate ion:

$$C_A = (R^\pm) + (R^-) + (Am^{--}) \quad (28)$$

Combining equations (26), (27), and (28) we obtain for the carbamate concentration

$$(Am^{--}) = C_A \left[1 + \frac{(H^+)}{(CO_2)K_{Am}'} \left(\frac{K_2' + (H^+)}{K_2'} \right) \right]^{-1} \quad (29)$$

If we start with a mixture of the amino acid and its sodium salt, and then expose the system to CO_2 at a partial pressure, p , the total dissolved CO_2 may be present in any one of the forms: CO_2 , H_2CO_3 , HCO_3^- , CO_3^{--} and Am^{--} . Ordinarily the concentrations of H^+ and OH^- are negligible in the equation of electrical neutrality, and the sodium ion is the only cation present in significant concentration in such a system. Then from the condition of electrical neutrality we have

$$(Na^+) = (R^-) + 2(Am^{--}) + (HCO_3^-) + 2(CO_3^{--}) \quad (30)$$

When the amino acid solution is exposed to CO_2 , however, the formation of Am^{--} proceeds far more rapidly than that of carbonate or bicarbonate, because of the slowness of the hydration of CO_2 to H_2CO_3 in the absence of a catalyst. This is particularly true if the reaction is carried out at low temperature, near 0° . It is thus experimentally possible to study the system, using rapid techniques, during a brief period when reaction (25) has come to equilibrium, and before any appreciable amount of bicarbonate or carbonate has formed. Under these circumstances, (30) simplifies to

$$(Na^+) = (R^-) + 2(Am^{--}) \quad (31)$$

This "noncarbonate" equilibrium has been studied particularly by Stadie and O'Brien (1935-36, 1937), with illuminating results. The total CO_2 in the liquid phase, in such a solution, is essentially equal to $(CO_2) + (Am^{--}) = g'p + (Am^{--})$, so that knowledge of p and of total CO_2 permits calculation of (Am^{--}) by difference.

From (27), (28), and (31), we obtain for (H^+)

$$(H^+) = K_2' \frac{(R^\pm)}{(R^-)} = K_2' \frac{[C_A - (Na^+) + (Am^{--})]}{(Na^+) - 2(Am^{--})} \quad (32)$$

This gives a relation between the experimentally measurable quantities (H^+) and (Am^{--}) which does not explicitly involve K_{Am}' . It does, of

course, assume the validity of the general reaction scheme as outlined, and holds only in the brief period during which the carbamate equilibrium has been established and the formation of bicarbonate and carbonate has not begun to any significant extent. Stadie and O'Brien showed that, during this period, (32) gave a good fit to the experimental measurements of $p\text{H}$ and (Am^{--}) . Moreover by combining (29) and (32) they obtained an expression from which (H^+) was eliminated:

$$(\text{Am}^{--}) \frac{[C_A - (\text{Na}^+) + (\text{Am}^{--})]}{[(\text{Na}^+) - 2(\text{Am}^{--})]^2} = (\text{CO}_2) \frac{K_{\text{Am}}'}{K_2'} = q'p \frac{K_{\text{Am}}'}{K_2'} \quad (33)$$

If the expression on the left-hand side—involving only quantities experimentally determined—is plotted against the partial pressure of CO_2 , a straight line should be obtained, of slope $q'K_{\text{Am}}'/K_2'$. This was found to be the case for solutions of glycine, alanine, and cysteic acid, in combination with their respective sodium salts; and since q' and K_2' are already known, K_{Am}' can be calculated immediately. Thus Stadie and O'Brien obtained values for pK_{Am}' at 20° of 5.5 to 5.6 for glycine and alanine, and 6.6 to 6.8 for cysteic acid.⁶ The higher value found for cysteic acid—which in the form that reacts with CO_2 may be written $-\text{O}_3\text{S}\cdot\text{CH}_2\cdot\text{CH}(\text{NH}_2)\cdot\text{COO}^-$ —is to be expected on electrostatic grounds. A new, negatively charged COO^- group has to be formed by the reaction of CO_2 with the amino group. In cysteic acid, electrostatic work must be done against the repulsion of two negatively charged groups already present; in the cases of glycine and alanine against only one such anionic group.

As p increases, (Am^{--}) rises in an approximately hyperbolic curve, approaching a limiting value at high p which is equal to $(\text{Na}^+)/2$, as may be seen from (31) if (R^-) approaches zero. For glycine and alanine at 20° , with equimolar mixtures of the amino acid and its sodium salt, (Am^{--}) attains about 80% of this limiting value at $p = 15$ to 20 mm Hg. For cysteic acid, under similar conditions, pressures several times as high are required to reach the same degree of carbamate formation, corresponding to the higher value of pK_{Am}' for this substance.

The transitory "noncarbonate" equilibrium, just described, is gradually replaced by a stable equilibrium in which bicarbonate and carbonate

⁶ Stadie and O'Brien report that pK_{Am}' decreases (K_{Am}' increases) with rising temperature. This would indicate that reaction (25) is endothermic (ΔH positive) when it proceeds from left to right. Pinsent *et al.* (1956b) studied the total heat evolved in the successive reactions: $\text{NH}_3 + \text{CO}_2 \rightarrow \text{H}_2\text{N}\cdot\text{COO}^- + \text{H}^+$ (1) and $\text{H}^+ + \text{NH}_3 \rightarrow \text{NH}_4^+$ (2). They found $\Delta H = -9.0$ kcal/mole at 0° , -12.7 at 20° and -15.6 at 40° . For reaction (2) it is known (see p. 452) that ΔH at 25° is -12.5 and $\Delta C_p \cong 0$. This gives ΔH for reaction (1) as $+3.5$ at 0° and -3.0 at 40° . Pinsent *et al.* (1956b) also determined the kinetics of reaction (1) above; the velocity constant $k' = -d(\text{CO}_2)/dt[(\text{CO}_2)(\text{NH}_3)]$ was $74 \text{ M}^{-1} \text{ sec}^{-1}$ at 0° and 1130 at 40° .

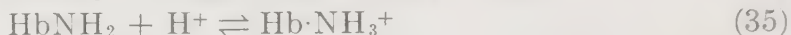
ions are present (equation 30). We shall not consider the details of this equilibrium here; they have been extensively discussed by Faurholt (1925) and by Stadie and O'Brien (1935-36). Equations (26) through (29) remain valid, and (31) is replaced by (30). Given a fixed initial ratio of the isoelectric amino acid to its sodium salt, before CO_2 is added, the main result is that the amount of carbamate formed rises rapidly from zero at $p = 0$, goes through a maximum at a low value of p (below 3 mm Hg), and then falls gradually to zero as p increases. Increase of p of course makes the solution more acid, hence decreasing (R^-) and increasing the amount of the dipolar ion (R^\pm), which is incapable of carbamate formation. At zero p , of course, no carbamate can be formed, and at high p the acidity of the solution represses carbamate formation, so the existence of a maximum is to be expected. The amount of carbamate formed at the maximum is of the order of 40% of the largest possible amount obtainable by converting all the R^- originally present into Am^{--} .

Since virtually all proteins contain some free amino groups— ϵ -amino groups of lysine, or terminal α -amino groups at the ends of peptide chains, or both—it may be expected that all proteins should form carbamino derivatives, at least in moderately alkaline solution. Few proteins have been studied in this connection, but the existence of carbamino compounds has been shown for hemoglobin and for the mixed proteins of blood plasma. The effects in hemoglobin are the most carefully studied and almost certainly the most important. Qualitatively, hemoglobin, in equilibrium with carbon dioxide, behaves like the amino acids. A "non-carbonate" equilibrium is rapidly attained, followed by a slower adjustment to a final equilibrium involving bicarbonate and carbonate ions.⁷ The amount of carbamate formation goes up as the negative net charge on the hemoglobin molecule increases, as might be expected. The maximum amount of bound carbamate, in the "noncarbonate" equilibrium, rises as high as 12 to 16 moles per mole of hemoglobin, when the negative net charge on the hemoglobin⁸ is -24 ; isoelectric hemoglobin was not found to bind any CO_2 as carbamate. The reaction can be formulated by essentially the same equation used to describe the combination in the case of the amino acids. Denoting hemoglobin, with one of its amino

⁷ Even "purified" hemoglobin solutions generally contain significant amounts of carbonic anhydrase, which accelerates the hydration of CO_2 so much that the two phases of CO_2 uptake cannot be separated. Carbonic anhydrase is readily poisoned, however, by addition of cyanide or sulfanilamide at low concentrations, and the addition of one of these substances to the hemoglobin permits the "noncarbonate" equilibrium to be readily studied.

⁸ We have expressed these figures per mole of hemoglobin (molecular weight 66,800); Stadie and O'Brien expressed them per mole of oxygen combining capacity (equivalent weight 16,700). Hence our figures are four times as large as those given by them.

groups, as HbNH_2 , we may write



Thus, as in all such systems, CO_2 and H^+ ions compete for the available amino groups. If the hemoglobin molecule contained n such amino groups, and if they were all equivalent and independent, then by the same argument used for the binding and release of protons in Chapter 9, equations (27) and (28),⁹ we could describe the whole system in terms of a single constant for the carbamate formation (K_{Am}') and a single constant (K_2') for proton dissociation from HbNH_3^+ . In that case, equation (29) would still apply to the situation, if we take (Am^{--}) as the total concentration of bound carbamate in *equivalents* per liter, C_A as the total number of loci on the hemoglobin molecules which are available for combination with CO_2 , also in equivalents per liter; and $(\text{Am}^{--})/C_A$ as the fractional saturation of these loci with CO_2 . If T_{Hb} is the total concentration of hemoglobin in all forms, in moles per liter, then $C_A = nT_{\text{Hb}}$. The ratio $(\text{Am}^{--})/T_{\text{Hb}} = \bar{\nu}$ is the average number of moles of CO_2 bound as carbamate per mole of hemoglobin, which may have any value between zero and n .

The number of free amino groups per molecule of horse hemoglobin is approximately 44. There are six terminal amino groups at the ends of peptide chains, all of which are valyl residues in horse hemoglobin. These should have a pK_A' value—corresponding to K_2' in equation (29)—near 7.8, on the basis of comparisons with pK_A' values of peptides (Chapter 8, Table VI). There are approximately thirty-eight ϵ -amino groups of lysyl residues (Chapter 3, Table II) which should have a pK_A' value near 10.5, if they are free to ionize.¹⁰ This would suggest that we ought to

⁹ See also the general discussion of interactions in Chapter 11. The situation treated here is a special case of one of the general classes of interactions discussed in that chapter.

¹⁰ There are indications (Steinhardt and Zaiser, 1955) that some of these lysine residues may not be free to ionize in the native protein. It is certain from the work of Steinhardt and Zaiser that there are a large number of acid and basic groups—probably thirty-six pairs of such groups—which are not free to ionize in the native protein, but become free on denaturation. There is considerable probability, but as yet no final proof, that many of the lysine residues are among these bonded (or blocked) groups in the native protein.

The recent work of Rhinesmith, Schroeder, and Pauling (1957) indicates that there are only four N-terminal α -amino groups in human hemoglobin, which fall into two different classes, although valine is the N-terminal residue in all four. It seems probable, in view of the findings in this work, that hemoglobin of other species also may have only four N-terminal residues—i.e. four peptide chains—and that earlier estimates were too high.

set up an equation for hemoglobin, analogous to equation (29), but with two terms on the right-hand side, each of the form there given, one for the six α -amino groups and one for the ϵ -amino groups, if we could treat all the groups of each class as equivalent and independent. We do not know, however, the appropriate values of K_{Am}' to assign for each of these two classes of groups. Since Stadie and O'Brien found values of $\bar{\nu}$ up to 16, when the mean net charge on the hemoglobin molecules was -24 , it is clear that other amino groups in hemoglobin besides the six α -amino groups can bind CO_2 as carbamate. These six, however, should be the first to react with CO_2 , because of their low value of K_2' , unless they also have very low values of K_{Am}' (see equation 29).

Making use of the relations $C_A = nT_{Hb}$ and $(Am^{--}) = \bar{\nu}T_{Hb}$ for hemoglobin, and assuming tentatively only one class of amino groups which combine with CO_2 , we may rewrite (29) in the form

$$\frac{\bar{\nu}}{n} = \frac{K_2'K_{Am}'(CO_2)}{K_2'K_{Am}'(CO_2) + K_2'(H^+) + (H^+)^2} \quad (36)$$

If we assume $K_2 \gg (H^+)$, this becomes

$$\frac{\bar{\nu}}{n} = \frac{K_{Am}'(CO_2)}{(H^+) + K_{Am}'(CO_2)} \quad \text{or} \quad \frac{\bar{\nu}}{n - \bar{\nu}} = \frac{K_{Am}'(CO_2)}{(H^+)} \quad (37)$$

and if $(H^+) \gg K_2'$, it becomes

$$\frac{\bar{\nu}}{n} = \frac{K_2'K_{Am}'(CO_2)}{(H^+)^2 + K_2'K_{Am}'(CO_2)} \quad \text{or} \quad \frac{\bar{\nu}}{n - \bar{\nu}} = \frac{K_2'K_{Am}'(CO_2)}{(H^+)^2} \quad (38)$$

Stadie and O'Brien found they could fit their data for both the "non-carbamate" and the total equilibria by a formula equivalent to (37), taking n as equal to the numerical value of \bar{Z} , the mean negative net charge on the protein. On this basis they obtained a value of pK_{Am}' equal to 6.5 to 6.8; but the approximations made in the use of (37) instead of (36) and the assumption that $n = |\bar{Z}|$ make the theoretical interpretation of their data uncertain. In any case the work of Ferguson and Roughton (1934), of Ferguson (1936), and of Stadie and O'Brien (1937) definitely proves the formations of carbamates of hemoglobin; and the latter authors have also shown the formation of carbamates of plasma proteins, although in much smaller amounts.

A significant fraction of the total carbon dioxide of the blood is transported in the form of carbamate under physiological conditions; according to Ferguson and Roughton this amounts to about 1 millimole of CO_2 bound as carbamate per liter of red blood corpuscles in arterial blood, and about 1.6 in venous blood. Stadie and O'Brien estimate somewhat

higher values for arterial blood—about 1.35 millimoles of CO_2 bound as carbamate per liter of whole arterial blood, of which they assign about one-third to the plasma proteins. They also give a figure for venous blood near 1.6, the increase above the value for arterial blood occurring almost entirely in the red cells. These figures are based on studies on horse and ox blood. Ferguson (1936) gives somewhat higher figures from direct measurements on human blood. All these investigators agree that venous blood has a higher content of carbamate than arterial blood, and attribute an important physiological function to carbamate in the transport of carbon dioxide from the tissues to the lungs, and its discharge in the lungs. Some 15 to 20% of the difference in total CO_2 between venous and arterial CO_2 is attributed by them to this difference in carbamate; Ferguson gives even higher figures. On the other hand, one of us (Wyman, 1948, pp. 485–496) has concluded that this difference is small, and of relatively minor physiological importance. The arguments involved are intricate and detailed; we shall not try to present them here, but merely call the attention of the reader to the existence of this difference of opinion, in case he wishes to examine the original papers.^{10a} In any case, there is no dispute concerning the presence of significant amounts of CO_2 as carbamate in blood; the evidence for carbamates of hemoglobin and of plasma proteins under physiological conditions appears to be decisive.

Rates of the Reactions $\text{CO}_2 + \text{H}_2\text{O} \rightleftharpoons \text{H}_2\text{CO}_3$

As has already been pointed out, the hydration of carbon dioxide and the dehydration of carbonic acid are processes that require a measurable time—a time so long as to be of crucial importance as a limiting factor in many biological processes. Therefore no discussion of the physical chemistry of carbon dioxide and its biochemical implications could be complete without a consideration of the rates of these processes.

By simple experiments it may readily be seen that these reactions require a finite time. The hydration reaction may be observed, for instance, by making up a cold solution of water saturated with CO_2 —it is convenient to do this by adding a little dry ice to the water—in the presence of a suitable indicator, such as bromothymol blue, which is yellow at pH 6, green near pH 7, and blue at pH 8. Such a solution is yellow if no alkali has been added. On addition of a little alkali (say 0.2 ml

^{10a} Roughton has not accepted Wyman's conclusions and has maintained the view that carbamino compounds play an important part in CO_2 transport in blood. See Roughton, F. J. W. (1949) *Biochem. J.* **44**, xxxi and *J. Physiol.* **109**, 12P; also Roughton's article in "Handbook of Respiratory Physiology," 1954, Air University, USAF School of Aviation Medicine, Randolph Field, Texas, p. 51 ff. (see especially pp. 86–88). We shall return to a more detailed discussion of this question in Volume II.

of 0.1 *N* NaOH to 5 ml of carbonic acid solution) the solution immediately becomes blue; in a second or two, the blue color fades, and the solution becomes yellow again. On addition of more alkali, the blue color immediately appears again, and then fades more slowly to green and yellow. The process can be repeated several times, until finally so much alkali has been added that the solution remains permanently blue.

The interpretation of these phenomena is as follows. The H_2CO_3 present immediately reacts with OH^- ion to form HCO_3^- and some CO_3^{--} . Most of the carbonic acid present, however, is in the anhydrous form of CO_2 . Before this can react, it must be hydrated, and this process takes an appreciable time. The H_2CO_3 so formed immediately reacts with the alkali; thus more H_2CO_3 continues to form, until all the alkali present is neutralized by it, or until all the CO_2 present is used up. The course of the process is seen most clearly in a cold solution, since the hydration reaction proceeds more slowly than at room temperature.

Similarly, by adding acid to a bicarbonate buffer, and observing *pH* changes with a suitable indicator, the process of dehydration of H_2CO_3 may be observed to take an appreciable time.

The first detailed study of these processes was made by Faurholt (1924). He concluded that in solution at *pH* < 8.0, if there is no exchange of CO_2 between the liquid and the gas phase during the process, the kinetics of the reaction may be expressed by the equation for the rate of change of concentration of dissolved CO_2 with time in the liquid phase.

$$\frac{d(\text{CO}_2)}{dt} = -k_{\text{CO}_2}(\text{CO}_2) + k_{\text{H}_2\text{CO}_3}(\text{H}_2\text{CO}_3) \quad (39)$$

That is, the CO_2 disappears by being hydrated at a rate proportional to its own concentration at any moment, while it is formed from H_2CO_3 at a rate proportional to the concentration of the latter. Since the concentration of the water in the system is large and essentially constant, the water concentration is implicitly included in k_{CO_2} . Both k_{CO_2} and $k_{\text{H}_2\text{CO}_3}$ are increased by many buffers—especially by the basic constituents of such buffers—as we shall see later. At *pH* values greater than 8.0, a second reaction must be taken into account, namely, $\text{CO}_2 + \text{OH}^- \rightleftharpoons \text{HCO}_3^-$; and the kinetic equation for this reaction may be written

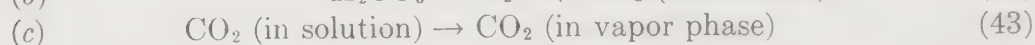
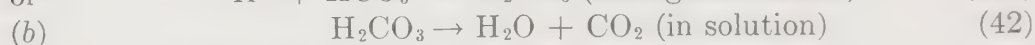
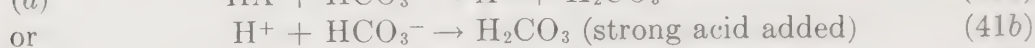
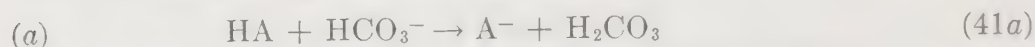
$$\frac{d(\text{CO}_2)}{dt} = -k_{\text{CO}_2, \text{OH}}(\text{CO}_2)(\text{OH}^-) + k_{\text{HCO}_3}(\text{HCO}_3^-) \quad (40)$$

At *pH* > 10, this second reaction is much faster than the first, but since we are considering only physiological systems, in which *pH* < 8, we shall need to consider only the first reaction (39). These velocity constants may be determined in various ways. We shall outline some of them here,

and refer briefly to the others; the reader seeking further information will do well to consult two important reviews by Roughton (1935, 1943-44), where many further references are also given.

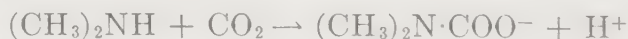
THE VELOCITY OF THE DEHYDRATION REACTION: $\text{H}_2\text{CO}_3 \rightarrow \text{H}_2\text{O} + \text{CO}_2$

A solution of bicarbonate is mixed with a suitable buffer solution, which we may designate by $\text{HA} + \text{A}^-$, or with a strong acid such as H^+Cl^- . The following reactions then occur:



Reaction (41a) or (41b) is very fast; it is never a limiting factor in determining the speed of the whole process. The speed of (43) depends on the extent of the surface which is exposed between the liquid and this vapor phase; it can be made very fast by shaking the liquid vigorously, so that it is constantly exposing new elements of surface to the vapor. Under these circumstances, (42) is the rate-determining reaction, and the process can be followed by measuring the increase of CO_2 in the vapor phase. This can be done by attaching a manometer to the reaction vessel; the increase of pressure is proportional to the amount of CO_2 evolved, and the proportionality factor can readily be determined by a suitable calibration.

On the other hand, reaction (43) may be suppressed by working in a vessel containing very little vapor, and by avoiding shaking. Faurholt (1924) worked under these conditions; he then determined the amount of (CO_2) in the liquid phase, at various time intervals, as follows. At the end of a given time interval, the reaction was suddenly stopped by throwing into the solution an excess of dimethylamine, followed by BaCl_2 . The dimethylamine reacts with CO_2 to form a carbamino compound, which is soluble.



The barium chloride precipitates all the HCO_3^- and CO_3^{--} present, as insoluble BaCO_3 . The BaCO_3 is removed by centrifuging, and the CO_2 combined with the dimethylamine can then be released by adding acid, and determined by any convenient method. The formation of the carbamino compound, and the precipitation of BaCO_3 , are both so rapid, at least at low temperatures, that there is not time for any appreciable interconversion of CO_2 and H_2CO_3 during the addition of the reagents.

As we have already indicated, it is also possible to measure the

progress of reaction (42) by determining the $p\text{H}$ of the system as a function of time. The rates are so fast, however, that most techniques cannot follow them with sufficient precision. The technique of Hartridge and Roughton for following the velocities of rapid reactions, which was first used for studying the velocity of the reactions of hemoglobin with oxygen and carbon monoxide, is a powerful tool in such cases. (For a general survey of methods of studying rapid reactions, see Roughton and Chance, 1953.) In principle it is very simple (Fig. 4). Two reservoirs (R_1 , R_2), containing the two solutions which are to be mixed together to produce

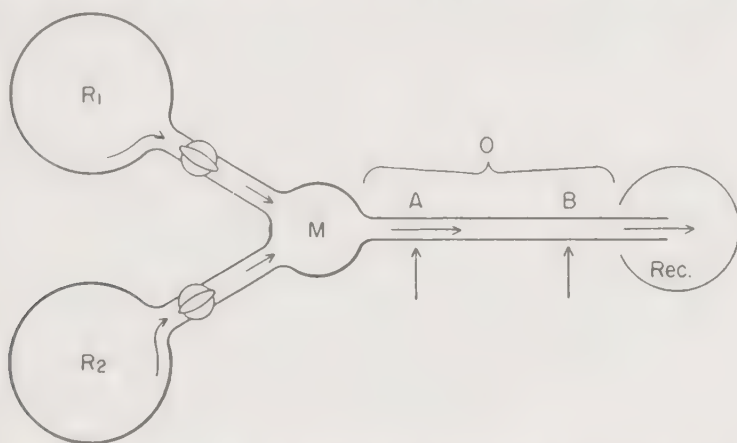


FIG. 4. Schematic diagram of Hartridge-Roughton apparatus for determining velocity of rapid reactions. R_1 and R_2 are reservoirs containing the reactant solutions. M is the mixing chamber; this contains a system of vanes and baffles, which disturb the flow so as to cause rapid mixing of the solutions from R_1 and R_2 . The observation tube O contains the turbulent flowing liquid, which discharges into the receptacle REC. Given the rate of discharge, and the cross-section of O , the velocity of the liquid in O is readily calculated. The distance between two observation points A and B thus gives the time taken by the liquid to move from A to B . Any physical method which gives the extent of the chemical reaction at different observation points thus determines the rate of the reaction.

the reaction under study, have outlet tubes which lead to the mixing chamber, M , where the liquids are mixed very rapidly by means of a set of inflow jets, so adjusted that mixing is virtually complete when the mixed fluid starts to flow down the observation tube, O . A constant pressure is maintained on the two reservoirs, so that the rate of outflow from the observation tube is constant. The time required for the liquid to move from M to any given point on O is readily calculated. If the outflow from O is $V \text{ cm}^3 \text{ sec}^{-1}$, and the cross section of O is $A \text{ cm}^2$, then the velocity of the fluid along the axis of O is $V/A \text{ cm sec}^{-1}$. The fluid in O is in a steady state, the extent of reaction at any point along the tube being independent of time. The time required for any given portion of fluid to

move a distance x along the tube, measured from M as starting point, is xA/V seconds. It is important that the flow of the liquid should be turbulent—not laminar streamline flow, such as occurs in a capillary viscometer—so that the whole body of liquid moves through O with essentially uniform velocity, whether it is near the center of O or near the wall. To obtain turbulent flow, the velocity must be fairly high, and the tube not too narrow. If these conditions are met, we may make observations of any suitable physical property of the solution, at different points along O , which can serve as an indicator of the extent of reaction at different time intervals after the reaction has begun in the mixing

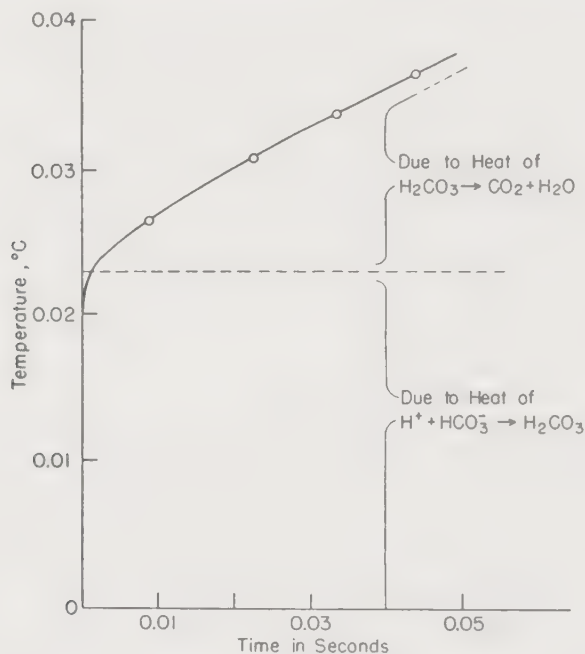


FIG. 5. Time course of heat evolution when hydrochloric acid is mixed with sodium bicarbonate, showing the two-stage character of the process. From Roughton (1941).

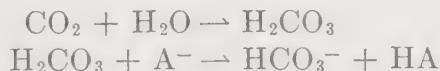
chamber. For instance, if the light absorption of the solution changes during the process, it can be readily measured at different points. Thus changes of pH with time can be determined in such an apparatus if a suitable indicator is present in the liquid. Alternatively, it is possible to seal in small electrodes at several points along O , and measure the pH electrometrically. The most precise measurements of reaction (42) by the rapid-flow technique, however, have been obtained by measuring the rise of temperature which accompanies the reaction, by means of thermocouples placed at suitable points along the tube, and thus determining the heat evolved as a function of the time (Roughton, 1941). When hydrochloric acid is mixed with sodium bicarbonate there is an essentially

instantaneous production of heat due to the formation of H_2CO_3 from H^+ and HCO_3^- (reaction 41b), followed by a slower evolution of heat due to the dehydration of H_2CO_3 (reaction 42) (Fig. 5). The total heat change involved in the two steps is indicated by the ΔH° values in Table IV, for $K_{\text{H}_2\text{CO}_3}$ and K_h respectively.¹¹ The rate of heat production in the dehydration process is thus proportional to $-d(\text{H}_2\text{CO}_3)/dt$ in reaction (39), and the velocity constant of this reaction, $k_{\text{H}_2\text{CO}_3}$, can thus be determined. The basic constituents of many buffers catalyze the hydration and dehydration reactions; hence if velocity constants are measured in a buffer solution, they should be determined at several different concentrations of buffer and then extrapolated to zero buffer concentration, as in the work of Kiese (1941).

THE HYDRATION REACTION: $\text{CO}_2 + \text{H}_2\text{O} \rightarrow \text{H}_2\text{CO}_3$

Here several types of procedure have been used. For instance we may shake a suitable buffer mixture with gaseous CO_2 , and follow the CO_2 uptake manometrically. This is, of course, the exact converse of the first method described by studying the dehydration reaction. The same reactions are involved, except that they go in the reverse direction.

Alternatively, we may mix a solution of dissolved CO_2 with a suitable buffer mixture ($\text{HA} + \text{A}^-$). The reactions are



and the first reaction is rate-determining. The change of (CO_2) with time may be followed by essentially the same methods used in studying the dehydration reaction.

Another elegant and entirely independent method for determining the rate of the hydration reaction was developed by Mills and Urey (1940). It involves the use of water containing an excess of the oxygen isotope, O^{18} , and the determination of the rate of exchange of this isotope between carbon dioxide and water. We shall not discuss the details of their calculations here, but their results for the velocity constants are listed, along with those of other workers, in Table VII (compare Rough-ton, 1943-44, p. 109).

At all temperatures studied, the velocity constant of the dehydration reaction is far greater than that of the hydration reaction, although the temperature coefficient of the latter is slightly greater. The ratio of the two velocity constants, when the system is at equilibrium, gives the equi-

¹¹ Note that the values of ΔH in Table IV indicate the heat *absorbed*, per mole, when the reaction runs in the opposite direction to that shown in Fig. 5.

librium constant, K_h (equation 4), as may readily be seen by setting $d(\text{CO}_2)/dt = 0$ in (39). The values of K_h calculated from the ratio of velocity constants by Roughton's thermal data (Table VII) diverge considerably from those given in Table III, which are derived from the ratio of the two equilibrium constants $K_{\text{H}_2\text{CO}_3}$ and K_1 . Thus from Table III we should conclude that CO_2 molecules outnumber H_2CO_3 molecules in aqueous solution at 0° by 500 to 1; from Table VII, by 950 to 1. The discrepancies at higher temperatures are even more pronounced, and there are marked disagreements at 25° among the different methods for determining the velocity constant, k_{CO_2} . Actually if we take Kiese's manometric value for k_{CO_2} , and the thermal value for $k_{\text{H}_2\text{CO}_3}$, at 25° , we obtain $K_h = 0.70/26.6 = 0.00263$, in very close agreement with the independent value of 0.00258 given in Table III; but no such close agreement is found if we compare the corresponding sets of values at 0° . All the measurements indicate that CO_2 is overwhelmingly predominant over H_2CO_3 at equilibrium, at all temperatures studied, and that the proportion of H_2CO_3 rises with rising temperature, the latter fact being a necessary consequence of the thermal data. Further research is needed, however, to resolve some of the secondary discrepancies indicated by all these studies.

The data of Table VII are for studies in unbuffered systems, or for buffered systems studied at several buffer concentrations and extrapolated to zero buffer concentration. Many buffers—specifically the basic components of these buffers—have marked catalytic effects on the hydration and dehydration reactions, as was shown particularly by Roughton and Booth (1938) and by Kiese and Hastings (1940). In the presence of buffers we may write (39) in the form

$$d(\text{CO}_2)/dt = -k_{\text{CO}_2}[1 + l(\text{B})](\text{CO}_2) + k_{\text{H}_2\text{CO}_3}[1 + l(\text{B})](\text{H}_2\text{CO}_3) \quad (44)$$

Here (B) denotes the molar concentration of the basic constituent of the buffer mixture, and l the catalytic coefficient of the base. Since the base acts as a catalyst, and does not alter the final equilibrium, it must increase the velocity constants of both the hydration and the dehydration reactions by the same factor, as is indicated in (44). Some values of l found were (1) for the anionic bases conjugate to weak uncharged acids: phosphate, 8; borate, 150; selenite, 2000; arsenite, 10,000; (2) for certain cyclic nitrogenous bases: imidazole, 1.5; 2,4-dimethylimidazole, 12; nicotine, 13 (Roughton and Booth, 1938). Bromine has a very powerful effect, as shown by Kiese and Hastings (1940); it surpasses any of the bases mentioned above, in its catalytic power. The catalytic effects in general are characteristic of what Brönsted has denoted as a generalized basic catalysis. None of the catalysts, however, compared in effectiveness with the enzyme carbonic anhydrase, discussed below, which on a

weight basis is more than a thousand times more effective than any of them.

These velocity constants may be employed in considering the rates of uptake of carbon dioxide by the blood in the tissues, and its release in

TABLE VII
VELOCITY CONSTANTS FOR THE HYDRATION OF CO_2 AND THE DEHYDRATION OF H_2CO_3 AT VARIOUS TEMPERATURES

Dehydration, $k_{\text{H}_2\text{CO}_3}$		Hydration, k_{CO_2}			
Temperature (°C)	Thermal	Thermal	Isotope exchange	Manometric	K_h , thermal
0	2.0	0.0021	0.00205	0.00205	0.00105
15	11.0	0.014	0.018	0.00128
25	26.6	0.037	0.0275	(0.070)	0.00139
38	89.0	0.131	0.10	(0.36)	0.00147

All values of $k_{\text{H}_2\text{CO}_3}$ and k_{CO_2} are in sec^{-1} . Thermal values from F. J. W. Roughton, *J. Am. Chem. Soc.* **63**, 2930 (1941). Isotope exchange values from E. A. Mills and H. C. Urey, *ibid.* **62**, 1019 (1940). Manometric values from M. Kiese, *Biochem. Z.* **307**, 207 (1941); see also M. Kiese and A. B. Hastings, *J. Biol. Chem.* **132**, 267 (1940). Kiese's values at 25° and 38° are obtained by extrapolation of his measurements at lower temperatures, assuming a linear relation between k_{CO_2} and $1/T$ —a relation found to hold very accurately between 0° and 22.5°. It should be noted that Kiese's measurements were made in phosphate buffer mixtures, at several concentrations of added buffer mixture for each temperature; he then extrapolated the value at each temperature to zero buffer concentration, to obtain the limiting rate in the absence of the catalytic effect produced by the buffer. Pinsent *et al.* (1956a) give revised thermal values of k_{CO_2} as 0.0112 at 15°, 0.0257 at 25°, and 0.0620 at 40°. These would of course give even lower values of K_h than those listed above.

Roughton [*Harvey Lect.* **39**, 96 (1943-44)] has pointed out that the values of $k_{\text{H}_2\text{CO}_3}$, given here by the thermal method, agree closely with the mean of several values determined earlier by less accurate methods, although the latter differed considerably among themselves.

Values of the equilibrium constant for hydration,

$$K_h = \frac{\text{H}_2\text{CO}_3}{\text{CO}_2} = \frac{k_{\text{CO}_2}}{k_{\text{H}_2\text{CO}_3}}$$

are calculated from the thermal data. Discrepancies with the other data, especially at 25°, are to be noted. Note also discrepancies with the values of K_h given in Table III which were determined by methods entirely independent of those given here. See discussion in text.

the lungs. In the tissues CO_2 produced by metabolic processes diffuses into the blood at a rate determined by the gradient in p_{CO_2} between blood and tissues. The amount, however, that could be taken up in the blood by simple solution as CO_2 , with a small trace of H_2CO_3 , is very small compared to the actual uptake. Most of the CO_2 to be taken up and

transported must be hydrated to H_2CO_3 . This then reacts with basic groups in the blood proteins, especially with imidazole groups of the histidine residues in hemoglobin and probably with some of the terminal α -amino groups at the ends of peptide chains, to give bicarbonate. If we denote these basic groups in the proteins by the symbol PB, then the reaction may be written



Most of the CO_2 flowing into the blood is thus converted into bicarbonate; but before reaction (45) can occur, the hydration of CO_2 must precede it, and this reaction, as may be seen from the values of k_{CO_2} in Table VII, takes an appreciable time in the absence of a catalyst. If we integrate the equation $d(\text{CO}_2)/dt = -k_{\text{CO}_2}(\text{CO}_2)$ from zero time up to time t , we obtain the result

$$\ln [a/(a - x)] = 2.303 \log [a/(a - x)] = k_{\text{CO}_2}t \quad (46)$$

where a is the initial concentration of CO_2 at zero time, and $a - x$ is the concentration after time t , in seconds. Thus the time when the reaction is half complete—that is, when $a - x = a/2$ —is given at 38° by the equation

$$t_{1/2} = 2.303 \log 2/k_{\text{CO}_2} = 0.69/0.131 = 5.3 \text{ sec at } 38^\circ \quad (47)$$

A given portion of blood, however, spends only a very short time in passing through a capillary, of the order of 1 to 5 seconds. If a proper job is to be done of unloading carbon dioxide from the tissues into the blood, the hydration reaction should be much more than half complete—preferably 99% complete at least—during this short period, since the diffusion processes involved in the transfer of CO_2 from tissue to blood also take a significant amount of time. Obviously, from the calculation shown in (47), no such result will be achieved unless a highly effective catalyst is present. A small amount of CO_2 is indeed directly taken up by the amino groups of proteins as carbamate, in a relatively rapid reaction, as we have already seen; but the conversion of CO_2 to bicarbonate by reactions (42) and (45) is the most important factor. It was a consideration of the difficulties involved here, especially by O. M. Henriques and F. J. W. Roughton, that led to the discovery of the enzyme carbonic anhydrase in the red cells by Meldrum and Roughton in 1932.

A serious problem also arises in connection with the discharge of CO_2 from the blood to the lungs. In the lung capillaries, reaction (45) must run from right to left, producing H_2CO_3 , which must in turn be dehydrated to CO_2 before it can pass off into the alveolar air. Reaction (45) proceeds with very great speed in either direction, but the speed of

dehydration is determined by the velocity constant $k_{\text{H}_2\text{CO}_3}$. Table VII shows that this constant is much larger than k_{CO_2} , and the reaction accordingly can be largely complete in a second or two, even in the absence of a catalyst. Detailed calculations by Henriques and by Roughton (see Roughton, 1935), however, have shown that the transfer of carbon dioxide from the lung vessels to the alveolar air would also be very incomplete, in the time taken by the blood to pass through the lung capillaries, if it were not for the presence of carbonic anhydrase.

Carbonic Anhydrase

The existence of a powerful catalyst in erythrocytes, promoting the hydration of CO_2 and the dehydration of H_2CO_3 , was first clearly proved by the work of Meldrum and Roughton in 1932.¹² The contents of hemolyzed red cells, even after the hemoglobin had been removed by denaturation and precipitation, for example by treatment with alcohol and chloroform, still retained the power of greatly accelerating these processes, whereas blood plasma was totally inactive in this respect. Purification of the catalyst is complicated by the presence of hemoglobin in the red cell in amounts which are vast in comparison with those of any other protein; carbonic anhydrase represents perhaps 1% of the total protein. The rather drastic methods of removing the hemoglobin by denaturation, as mentioned above, have been employed by almost all investigators. Carbonic anhydrase is a remarkably stable protein and seems to survive this treatment to give a pure and highly active product; but it remains possible that the enzyme has undergone some alteration from its native state in the red cell in this initial step. It is not easily salted out, being precipitated by ammonium sulfate only at very high concentrations (85 to 100% saturation with $(\text{NH}_4)_2\text{SO}_4$). Thus most other proteins remaining in the red cell extract, after removal of hemoglobin, can be removed by precipitation with ammonium sulfate at concentrations up to 70% saturation, and the carbonic anhydrase then separated by precipitation in nearly saturated ammonium sulfate. Other techniques have also been used successfully in the purification process, but it is not our purpose here to describe such methods in detail.

The best preparations of carbonic anhydrase obtained by Roughton, by Kiese and Hastings, by van Goor, by Keilin and Mann, by Scott and Fisher, and by some other workers, are of extremely high activity. They have been reported to double the rates of hydration of CO_2 , or

¹² No attempt will be made to give detailed references here to the numerous papers on carbonic anhydrase. The reader who wishes to find them should consult the valuable reviews by Roughton (1935, 1943-44), van Goor (1948), Roughton and Clark (1951), and Vallee (1955).

dehydration of H_2CO_3 , at concentrations of one part in 100,000,000 or even less. Nevertheless it is by no means certain that any completely pure preparation of carbonic anhydrase has yet been achieved. Observations in the ultracentrifuge, or by electrophoretic methods, indicate impurities up to 15% in some of the best preparations. The molecular weight is of the order of 30,000, but the exact value is still somewhat in doubt; the isoelectric point is near 5.6. The most interesting and distinctive chemical property of the enzyme, however, is the fact that it contains zinc, as was first shown by Keilin and Mann, and confirmed by Scott and Fisher. The figures for zinc content given by different investigators differ somewhat, ranging from 0.20 to 0.33%, but the amount almost certainly corresponds to one atom of zinc per molecule of carbonic anhydrase. The zinc is indispensable for enzyme activity, and it is very tightly bound. Tupper, Watts, and Wormall (1952) prepared a solution containing carbonic anhydrase and a zinc salt containing Zn^{65} , and found no exchange of the isotopic zinc with the nonradioactive zinc in the enzyme over a period of 32 days.

Carbonic anhydrase is poisoned or inhibited by a number of substances. Cyanide inhibits the activity by 85% in a concentration as low as $4 \times 10^{-6} M$, according to Keilin and Mann. Azide is a strong inhibitor, and so are several metallic ions, including zinc itself, also Ag, Au, Cu, and Hg. Thiocyanate ions exert a marked inhibitory effect. Sulfanilamide (*p*-aminobenzenesulfonamide) is a powerful inhibitor, and a particularly valuable one for studies with carbonic anhydrase, since it is not inhibitory for most other enzymes. It can therefore be used in complicated systems to inhibit carbonic anhydrase specifically while leaving other enzymes in the system still functioning. The free $-\text{SO}_2\text{NH}_2$ group of sulfanilamide is apparently essential for this inhibitory action; if one of the hydrogens of this group is replaced by an alkyl group, the inhibitory action disappears.

Carbonic anhydrase is found in other tissues besides the red cells of the blood, notably in the pancreas, in which large amounts of CO_2 must be converted into bicarbonate in the pancreatic secretion. It is also present, in rather smaller amounts, in gastric mucosa, in the kidney, and in the salivary glands and retina. It is apparently completely absent in muscle, and in most other tissues of the body. The biological reason for its absence seems fairly clear, if it is CO_2 , not H_2CO_3 or bicarbonate ion, which is produced as the end product of metabolic activity. In the presence of carbonic anhydrase, the CO_2 would be rapidly hydrated to H_2CO_3 , and this would be converted to bicarbonate ion by the basic groups present in the tissue proteins or other tissue buffers. This would convert these buffer bases to their conjugate acids, with perhaps a dangerous

decrease in the buffer capacity of the tissue systems. Moreover, an even more serious trouble would be created, for carbonic acid and the bicarbonate ion diffuse out through the tissue cells and into the blood much more slowly than CO_2 . The absence of carbonic anhydrase in most tissue cells therefore favors the rapid outward passage of metabolic CO_2 . As Roughton (1935, p. 262) has well said: "In such a location carbonic anhydrase would be an enemy to the organism, rather than a friend."

For similar reasons, there are biological advantages in having the carbonic anhydrase of blood located entirely in the red cells, not in the plasma. This means that CO_2 , arriving from the tissues, passes rapidly through the plasma and enters the cells, where it is rapidly hydrated, and then converted to HCO_3^- ion by hemoglobin, the most plentiful and most powerful buffer of the blood. Much of the HCO_3^- ion so formed passes back into the plasma, by exchange with chloride ions from plasma across the red cell membrane—a process discussed in detail later in Volume II—but the work of neutralizing the acid has first been done in the red cells by the hemoglobin. If carbonic anhydrase were present in plasma, large amounts of H_2CO_3 would be formed there, before the CO_2 from which it is derived had time to diffuse into the red cells. Then this H_2CO_3 would be neutralized by the relatively weak buffering power of the proteins of plasma, and the resulting change of pH would, at least for a transitory period, be much greater than that which actually occurs in plasma while blood is passing through the tissues.

The cycle of events taking place in the lungs is of course the converse of that in the tissues, and the nature of the whole complex system involved is considered in detail in Volume II. Here we have sought only to indicate briefly the biological significance of carbonic anhydrase, and the value to the organism of its absence in some places as well as of its presence in others.

GENERAL REFERENCES

- Bassham, J. A., and Calvin, M. (1956). Photosynthesis, in "Currents in Biochemical Research 1956" (D. E. Green, ed.), p. 29. Interscience, New York.
- Calvin, M. (1956). The Photosynthetic Carbon Cycle, in "Proceedings of the Third International Congress of Biochemistry, Brussels 1955" (C. Liébecq, ed.), p. 211. Academic Press, New York.
- Henderson, L. J. (1913). "The Fitness of the Environment," The Macmillan Company, New York.
- Henderson, L. J. (1928). "Blood: A Study in General Physiology," Yale University Press, New Haven, Connecticut.
- Krebs, H. A., and Kornberg, H. L. (1957). A survey of the energy transformations in living matter. *Ergeb. Physiol. biol. Chem. u. Exptl. Pharmacol.* **49**, 212.
- Peters, J. P., and Van Slyke, D. D. (1932). "Quantitative Clinical Chemistry," Vol. I, 1st ed., Chapter XVIII.

- Rankama, K., and Sahama, T. G. (1950). "Geochemistry," Chapter 19, University of Chicago Press, Chicago.
- Roughton, F. J. W. (1935). Recent work on carbon dioxide transport by the blood, *Physiol. Rev.* **15**, 241.
- Roughton, F. J. W. (1943-44). Some recent work on the chemistry of carbon dioxide transport by the blood, *Harvey Lectures* **39**, 96.
- Roughton, F. J. W., and Chance, B. (1953). Rapid Reactions, in "Technique of Organic Chemistry" (A. Weissberger, ed.), Vol. VIII, Chapter X, p. 669, Interscience Publishers, New York.
- Roughton, F. J. W., and Clark, A. M. (1951). Carbonic Anhydrase, in "The Enzymes" (J. B. Sumner and K. Myrback, eds.), Vol. I, Part 2, p. 1250, Academic Press, New York.
- Vallee, B. L. (1955). Zinc and metalloproteins, *Advances in Protein Chem.* **10**, 317.
- Van Goor, H. (1948). Carbonic anhydrase: Its properties, distribution, and significance for carbon dioxide transport, *Enzymologia* **13**, 73.

SPECIAL REFERENCES

(see also references given in the tables)

- Berg, D., and Patterson, A., Jr. (1953). *J. Am. Chem. Soc.* **75**, 5197. The work reported in this paper has been corrected and extended by that of Wissbrun *et al.* (see Table III).
- Cohn, E. J., Green, A. A., and Blanchard, M. H. (1937). *J. Am. Chem. Soc.* **59**, 509.
- Faurholt, C. (1924). *J. chim. phys.* **21**, 400.
- Faurholt, C. (1925). *J. chim. phys.* **22**, 1.
- Ferguson, J. K. W. (1936). *J. Physiol.* **88**, 40.
- Ferguson, J. K. W., and Roughton, F. J. W. (1934). *J. Physiol.* **83**, 68, 87.
- Kiese, M. (1941). *Biochem. Z.* **307**, 207.
- Kiese, M., and Hastings, A. B. (1940). *J. Biol. Chem.* **132**, 267.
- Pauling, L. (1940). "The Nature of the Chemical Bond," 2nd ed., Cornell University Press, Ithaca, New York.
- Pinsent, B. R. W., Pearson, L., and Roughton, F. J. W. (1956a). *Trans. Faraday Soc.* **52**, 1512.
- Pinsent, B. R. W., Pearson, L., and Roughton, F. J. W. (1956b). *Trans. Faraday Soc.* **52**, 1594.
- Rhinesmith, H. S., Schroeder, W. A., and Pauling, L. (1957). *J. Am. Chem. Soc.* **79**, 4682.
- Roughton, F. J. W. (1941). *J. Am. Chem. Soc.* **63**, 2930.
- Roughton, F. J. W., and Booth, V. H. (1938). *Biochem. J.* **32**, 2049.
- Sanger, F. (1952). *Advances in Protein Chem.* **7**, 1.
- Stadie, W. C., and O'Brien, H. (1935-36). *J. Biol. Chem.* **112**, 723.
- Stadie, W. C., and O'Brien, H. (1937). *J. Biol. Chem.* **117**, 429.
- Steinhardt, J., and Zaiser, E. M. (1955). *Advances in Protein Chem.* **10**, 151.
- Tupper, R., Watts, R. W. E., and Wormall, A. (1952). *Biochem. J.* **50**, 429.
- Wyman, J. (1948). *Advances in Protein Chem.* **4**, 407.

Chapter 11

Some General Aspects of Molecular Interactions

In the discussion of acid-base equilibria, we have dealt with an important class of molecular interactions which are describable in terms of the law of mass action. This law is also the basis for describing a vast number of reactions which can and do proceed in the living organism or in extracts from living tissues which contain the necessary enzyme systems. In many cases one of the interacting molecules can bind the other at any one of a number of different binding sites. This is notably true for acid-base equilibria. Large protein and nucleic acid molecules may contain hundreds or thousands of basic sites to which protons can be attached. Other interactions of the sort to which we are referring here are continually encountered in biochemical systems of the most diverse sorts. An outstanding example is the combination of hemoglobin with oxygen; the four iron atoms of the hemoglobin molecule can each combine with one oxygen molecule, and the binding of an oxygen at one of the four positions markedly affects the affinity for oxygen of the unbound iron atoms at the other three positions. Another important example is the remarkable capacity of the molecule of serum albumin to combine with any one of a great number of substances; most of these are anions, but some are neutral molecules or cations. These include numerous anionic dyes, fatty acid anions, and nearly all simple inorganic anions, including the halogens, the nitrate ion, perchlorate ion, and many others. Indeed it is difficult to mention an anion which is not bound by serum albumin. Albumin also binds significant amounts of many neutral molecules, such as decanol and other higher alcohols, chloroform, and even benzene. Steroids, metalloporphyrins, and many other compounds are bound by serum albumin in appreciable amounts. One albumin molecule may bind as many as forty chloride or thiocyanate ions.

All proteins are capable of binding metallic cations, but the extent of binding varies greatly, both with the nature of the cation and with the nature of the protein and the net charge which it carries. The alkali metals are scarcely bound at all; the alkaline earths are bound in significant amounts by proteins, but in general only when the protein is negatively charged. The nucleic acids, which are of course negatively charged at all *pH* values except in strongly acid solution, likewise bind the alka-

line earths strongly. Much firmer attachments, to proteins at least, are made by the transition group elements, those of physiological importance being particularly manganese, iron, copper, and zinc—also molybdenum, in the light of recent studies on some of the metalloflavin enzymes. These, unlike the alkaline earths, which form bonds which are predominantly ionic in character, may also form bonds which are largely covalent. These complexes often exist in a mobile equilibrium with the molecule that binds the metal ions so that the combination can be described by one or more definite association constants; and it is such cases as this with which we shall be predominantly concerned here. In many substances of great biological importance, however, the binding of the metal to the protein is very tight, so that no dissociation of the complex is detectable at all. This is substantially true of hemoglobin in the physiological pH range; association constants for hemoglobin with oxygen or carbon monoxide are readily determined; but the dissociation of iron or of the iron-porphyrin complex from the hemoglobin molecule, at ordinary temperatures and at pH values between (say) 6 and 9, is undetectably small. Likewise the zinc atom of carbonic anhydrase is so firmly bound that it does not detectably exchange with isotopic zinc in the surrounding solution in a period of more than a month. (See Chapter 10.) There are a number of metalloenzymes for which similar statements can be made; the problem in such cases—a problem not yet fully solved for any case—is the determination of the chemical structure of the complex and its significance for the specific activity of the enzyme. This lies outside the scope of our discussion here. There are, however, plenty of cases in which activation or inhibition of enzyme action is produced by the readily reversible binding of small amounts of certain metallic ions, and such cases clearly fall within the domain of the problems to be considered in this chapter.

Complexes between enzymes and coenzymes such as the pyridine nucleotides are of great physiological importance for enzyme action, and the association constants involved are clearly definable in many cases. The formation of enzyme-substrate complexes themselves represents another branch of inquiry in the same field, although the kinetic problems that arise here require a further analysis which is outside the scope of this chapter.

The great class of oxidation-reduction reactions involve interactions of the most important sort. These are so important, and involve such special problems, that they will be considered in a later volume; but many of the general principles to be considered here are basic for the discussion of oxidation-reduction processes.

In addition to all these, there are interactions of great importance which involve linkages between two or more macromolecules. The spe-

cific combinations between antigens and antibodies, leading sometimes to the formation of soluble complexes of high molecular weight, sometimes to the formation of precipitates or to agglutination reactions, furnish an example of profound biological importance. (For discussion and references, see, for instance, Boyd, 1956.) The formation of nucleoproteins from proteins and nucleic acids is another process of major importance, but one which is still very incompletely understood. Some important proteins, such as insulin, exist in several different polymeric forms. The fundamental molecule of insulin, as we have observed in Chapter 3, is a unit containing 51 amino acid residues, with a molecular weight of 5734, for beef insulin, and practically the same for hog and sheep insulins, but it associates very readily to form a dimer of this unit—so strongly, indeed, that the monomer is almost undetectable in ordinary aqueous solutions. The dimer in turn associates to larger polymers; in neutral aqueous solutions the predominant species of insulin molecules are those which are six or eight times the size of the monomer unit (molecular weight 34,000 to 46,000). In acid solutions, particularly in the presence of divalent anions and at relatively high ionic strength, very extensive polymerization of insulin into fibrillar structures occurs. Other proteins, such as the hemocyanins, can undergo reversible association and dissociation, depending on changes in ionic strength. No attempt at detailed discussion of these matters will be made in this chapter, although some of them are dealt with later in the chapters on macromolecules and their properties. An extensive and valuable discussion of protein-protein interactions is given by Waugh (1954).

The interactions discussed here involve actual molecular complexes, the formation of which can be described in terms of the law of mass action. The bonds holding the constituents of the complex together may be very different in different cases, ranging from ion pairs which stick together essentially because of electrostatic attraction, to compounds held by strong covalent bonds. These interactions are to be distinguished from those due to ion-ion and ion-dipole attractions, such as those discussed in Chapter 5. The latter, for molecules and ions of high net charge or high dipole moment, may involve profound changes in the activity coefficients of the components which interact, as demonstrated by the far-reaching solubility changes associated with the salting in and salting out of proteins by neutral salts. In such a case as the salting in of β -lactoglobulin by sodium chloride, for example, there is practically no evidence of the binding of sodium or chloride ions by the protein. Yet its solubility is increased more than a hundredfold by an increase of sodium chloride concentration from zero to 0.1 *M*. On the other hand, the addition of a salt such as zinc or cadmium chloride to a protein solution certainly leads

to actual complex formation between the cation and the protein, and for some proteins such as serum albumin to complex formation with the anion as well.

The molecular complexes to be discussed here are usually formed from constituents some of which are ions; in many cases all the reactants, and the products as well, are ionic. In such cases variation of interionic forces, due for instance to variation of the dielectric constant or ionic strength of the medium, may have a profound effect on the observed equilibria. An added salt may have a relatively nonspecific effect, owing to the change of ionic strength which its addition produces, and a more specific effect, owing to the fact that one or both of its ions can be bound to a molecule which is present. In formulating the fundamental equilibrium relations later in this chapter, we shall consider first the cases where interionic forces can be neglected, at least as a first approximation, and deal later with cases where corrections for interionic forces are of major importance.

Some Methods of Measuring Binding

When two molecules interact, the determination of the amount of one bound by the other can be carried out in many different ways. The systems that may be studied are so diverse, and the methods of studying them so many, that it would be a futile task to attempt to list them all. Certain methods, however, are of such general importance that they deserve some brief discussion; but we shall attempt no detailed description of procedure. We shall speak of seven such methods: equilibrium dialysis, ultrafiltration, spectroscopic measurements, electromotive force measurements, ultracentrifuge measurements, electrophoretic mobility studies, and *pH* changes associated with binding.

EQUILIBRIUM DIALYSIS

In this procedure, a solution containing a macromolecule, *P*, is separated by a semipermeable membrane from another solution containing only the solvent and other small molecules or ions which can pass through the membrane. In the simplest case, the outer solution contains only solvent and the "ligand" molecule or ion, *A*, the binding of which to the macromolecule, *P*, is to be studied. The two solutions are stirred and remain in contact through the membrane until analysis shows that they have come to equilibrium. The simplest case arises if *A* carries no electric charge, and if its activity coefficient is independent of concentration. Then at equilibrium an analysis is carried out for the *total* concentration of *A*, both inside and outside the membrane. We denote the latter as $(A)_i$, the former as $(A)_o$. Then, if we assume that the macromolecule does

not affect the activity of A except by binding it, we have the relation

$$(A)_i = (A)_o + (A)_{\text{bound}} = (A)_o + \bar{\nu}(P) \quad (1)$$

where (P) is the molar concentration of the macromolecule and $\bar{\nu}$ is the mean number of molecules of A bound by one macromolecule under the conditions of the experiment. It is well to repeat the experiment for several values of (P) at constant $(A)_o$, in order to show that $\bar{\nu}$ is a well-defined number, independent of (P), at least within the limits of concentration under study. Variation of $(A)_o$ at constant (P) should of course give varying values of $\bar{\nu}$.

When A is charged, its binding to P necessarily alters the net charge on P; and if P carries a net charge, the distribution of diffusible ions, including those unbound, is necessarily different inside and outside the membrane, according to the laws of the Gibbs-Donnan equilibrium (Volume II). The correction required to take account of this effect can be calculated, or the effect itself can be reduced to very small proportions by working in a solution containing a considerable excess of a suitable neutral salt, as is shown in detail by the discussion of the Gibbs-Donnan equilibrium. There may be pitfalls in the latter technique, however; in the study of the binding of anions by serum albumin, for instance, the anion of any neutral salt employed is certain to be bound to some extent by the albumin; and in doing so it may compete for some of the same sites on the albumin molecule at which binding of the substance under study also occurs. Thus the addition of neutral salts or buffer mixtures may seriously complicate the interpretation of the measurements. Nevertheless the method of equilibrium dialysis is a very powerful tool and has provided many of the most important data in the study of interactions.

ULTRAFILTRATION

This is in principle very similar to equilibrium dialysis. The outer phase, free of macromolecule, is obtained by filtering some of the inner solution under pressure through a semipermeable membrane, and the inner and outer fluids are then analyzed as before. The filtration process itself somewhat alters the composition of the inner fluid, and this factor must be allowed for in calculating the results. This difference between the composition of the inner and outer fluids occurs when the macromolecule carries a net charge and is determined by the Donnan equilibrium. The correction factor is small, however, and may be negligible, if the volume of fluid filtered through the membrane is small compared to the amount inside.

SPECTROSCOPIC MEASUREMENTS

If either P or A absorbs light of some conveniently measureable wavelength, the absorption may be altered when they combine. An example of the change in light absorption by a ligand is shown in Fig. 1, which is taken from the work of I. M. Klotz on the interaction of azosulfathiazole with bovine serum albumin. The marked decrease in light absorption, in the neighborhood of 5000 Å, when the dye is bound to albumin, is apparent. The best-known case of this sort, however, is the striking change in

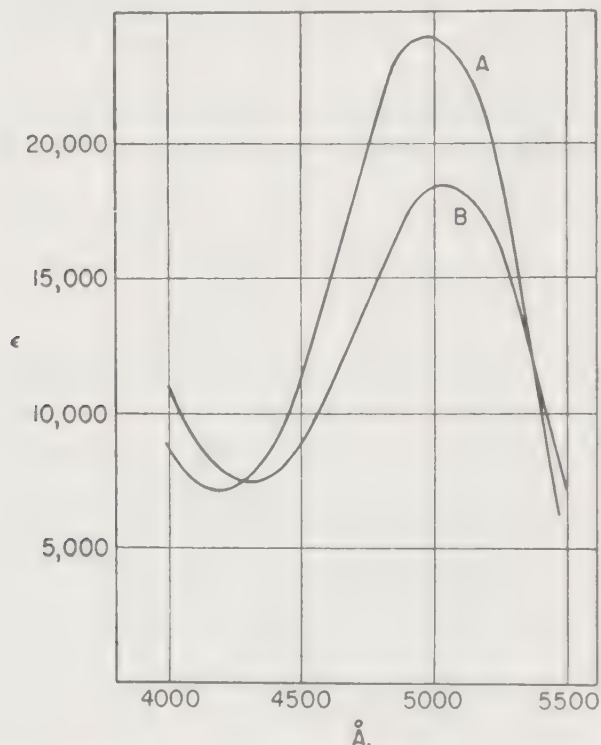


FIG. 1. Absorption spectra of azosulfathiazole, (A) in buffer at pH 6.9; (B) in buffer containing 0.2% serum albumin at pH 6.9. (From Klotz, 1953.)

the absorption bands of hemoglobin when it combines with oxygen or carbon monoxide. In this case the extinction coefficient, for any suitably chosen absorption band, appears to be a linear function of $\bar{\nu}$ —the value of $\bar{\nu}$ here varies between 0 and 4, since there are four binding sites on the hemoglobin molecule for O_2 or CO . The use of changes in absorption spectra as a means of studying the binding of protons to bases has already been discussed in some detail in Chapter 8 (pp. 424–428).

ELECTROMOTIVE FORCE MEASUREMENTS

In the particularly important case of proton binding, the use of the hydrogen or glass electrode to determine affinity constants has already

been discussed in Chapter 8. The same type of measurement can be applied to determining the activity of other ions in solution, if a suitable electrode, reversible to the ion in question, is available. Thus Scatchard *et al.* (1950) used electrodes of the type



Solutions I and II both contained sodium chloride at the same total concentration, in moles per kilogram of water, but solution II contained in addition human serum albumin. If no albumin were present, solutions I and II would be identical, the cell would be completely symmetrical, and the electromotive force of the cell would be zero, since every potential difference at a phase boundary on the left-hand side of the cell is balanced by an equal and opposite potential difference on the right-hand side. (This point was tested experimentally in the cells employed, to make sure that the pair of silver-silver chloride electrodes used actually were equivalent.) When albumin is present in solution II some chloride ion is bound, and the activity of chloride ion is lower in solution II than in solution I. By the same line of reasoning employed in the interpretation of measurements on the hydrogen electrode (Chapter 8, pp. 429-437), the electromotive force of cell (2) may be written, with primes to denote the quantities in the albumin-free solution:

$$E = \frac{RT}{F} \ln \frac{(\text{Cl}^-)'\gamma'}{(\text{Cl}^-)\gamma} = 0.0591 \log \frac{(\text{Cl}^-)'\gamma'}{(\text{Cl}^-)\gamma} \text{ at } 25^\circ \quad (3)$$

Here (Cl^-) and $(\text{Cl}^-)'$ denote the concentrations of *free* chloride ion in solutions I and II, respectively. The free chloride in solution I is of course the same as the total sodium chloride concentration. The *total* chloride is the same in II as in I, but the free chloride is less because some is bound by the protein. Likewise γ and γ' denote the activity coefficients of the free chloride ion in the two solutions. It is generally assumed that γ and γ' are the same; this assumption cannot be exactly true, but any error involved in it is likely to involve only a second-order correction, in a solution containing a substance which binds chloride ions as strongly as serum albumin. The concept of individual ion activities and the difficulties associated with the existence of liquid junction potentials have been discussed at length in Chapter 8, in connection with measurements on the hydrogen electrode, and the value and the limitations of the concept are the same for the activity of the chloride ion in cell (2) as for the hydrogen ion in the cell defined by (65) in Chapter 8. Given the knowledge of $(\text{Cl}^-)'$, which can be determined by direct analysis, the assumption that $\gamma = \gamma'$, and the measured value of E in (3), the concentration of unbound chloride ion in the solution containing the protein is imme-

diately obtained. The difference between this and the total chloride ion concentration in this solution (m_3), divided by the molal concentration of protein, m_2 , then gives

$$\bar{\nu} = \frac{m_3 - (\text{Cl}^-)}{m_2} \quad (4)$$

Scatchard *et al.* (1950) also carried out determinations of binding by the equilibrium dialysis method; the agreement of the values of $\bar{\nu}$ obtained by the two methods under the same conditions furnished strong evidence for the validity of both techniques. The results also indicated that, although chloride ion was strongly bound by albumin, there was no evidence for binding of sodium ion. They also studied the binding of thiocyanate ion by albumin, but the use of silver-silver thiocyanate electrodes involved special difficulties and precautions which we shall not discuss here. There are unfortunately many ions for which no suitable reversible electrode is available, and for which therefore the electromotive force technique cannot be applied.¹ The range of ion concentrations over which a reversible electrode gives trustworthy readings is much more limited for most electrodes than it is for the hydrogen electrode. Scatchard *et al.*, for instance, found that cell (2) behaved erratically at chloride ion concentrations below about $5 \times 10^{-4} m$.

ULTRACENTRIFUGE MEASUREMENTS

If the macromolecule, P, binds the small molecule, A, then P will carry A with it if it is caused to sediment in the ultracentrifuge. With a suitable device, the sedimenting cell can be divided into sections after sedimentation has proceeded for a given time, and the solution analyzed for P and A at different levels in the cell. A classical example is the work of Chanutin *et al.* (1942) on the binding of calcium ions by casein. In any typical experiment an analysis for casein and for (Ca^{++}) at different levels in the ultracentrifuge cell showed that the relation between the two is linear, and the extrapolated value of (Ca^{++}) at zero casein concentration gives the value of free calcium; with the total amount of added calcium known, the bound calcium can be calculated by difference. Some sedimentation of unbound (Ca^{++}) must obviously occur also, but the small concentration gradient of free Ca^{++} in the cell can fairly readily be corrected for. Recently this method has been used with important results by Velick and his associates (see, for instance, Hayes and Velick, 1954) to study the binding of coenzymes to enzymes such as glyceraldehyde-3-phosphate dehydrogenase and alcohol dehydrogenase of yeast and muscle. We shall consider these studies in more detail later.

¹ Consider however the use of ion exchanger electrodes; see pp. 650-651.

ELECTROPHORESIS

If a small ion is bound to a macromolecule, the net charge of the latter is altered, and correspondingly its mobility in an electric field must be altered also. If the protein is isoelectric—that is, the mobility is zero—in a certain buffer, and a neutral salt is then added to the solution, selective binding of the anion or cation of the salt is reflected by a change in the mobility to positive or negative values. If both anion and cation are equally bound, or if neither is bound, the protein remains isoelectric. Thus Longsworth and Jacobsen (1949) found the isoelectric point of bovine serum albumin to be at pH 4.71 in an acetate buffer containing 0.1 N sodium acetate. In a buffer containing 0.02 N acetate and 0.08 N sodium chloride the isoelectric point shifted to pH 4.59. In 0.01 N acetate plus 0.09 N sodium thiocyanate, it shifted to pH 4.17. The results indicate that chloride is bound more strongly than acetate ion (from other types of measurements it is known that acetate is also bound), and thiocyanate is bound much more strongly than either. Obviously if one of these anions is bound to a protein that was originally isoelectric in the absence of any added electrolyte, the protein can be made isoelectric once more by the binding of protons. At the new isoelectric point, the number of bound anionic charges must equal the number of bound protons. Conversely a protein such as myosin, which has a strong tendency to bind cations such as calcium, has its isoelectric point shifted to a higher pH value in a solution containing calcium ions than in one in which calcium and other divalent cations are absent.

The protein need not be isoelectric to apply this test. A change of mobility produced by the substitution of one ion for another, in a solution of any given ionic strength, is an index of some difference in the binding of one ion or the other, whether the original mobility was positive, zero, or negative. The factors influencing mobility of ions, however, are numerous and complex, so that it is more difficult to derive quantitative estimates of amounts bound by this method than by some of the other methods already described. (For further discussion and references, see Alberty, 1953.)

EFFECTS OF pH CHANGES WHEN SALTS ARE ADDED: ISOIONIC POINTS

Electrostatic effects of adding salt to a polyvalent acid or ampholyte, as they affect the binding or release of protons, have already been considered at length in Chapter 9, pp. 512–547. In addition, however, there are effects on the pH of such solutions when ions other than protons are bound.

Experiments have been carried out by first removing all small dif-

fusible ions as completely as possible, except for hydrogen and hydroxyl ions, which are obviously not removable in aqueous solutions.¹ If the ampholyte under study is present at not too low a concentration, the pH of the solution will correspond closely to the pH of the isoelectric point, defined as the point of zero mobility in an electric field, under the same conditions. The calculation of the pH of a salt-free solution of glycine at various concentrations has already been carried out in Chapter 9 (p. 507); the same general reasoning, with appropriate modifications, can be applied to more complex ampholytes. Addition of a neutral salt to such a solution may lead to some small pH changes due to interionic forces; but if one of the ions of the salt is selectively bound by the ampholyte, then the pH will generally change much more than would be expected from interionic forces alone.

It is easiest to visualize what happens in terms of a specific case. Scatchard and Black (1949) studied the effects of adding various neutral salts to a salt-free solution of human serum albumin. At a concentration of 7 g/l the concentration of albumin is very nearly $10^{-4} M$; the pH is near 5, so $(H^+) = 10^{-5}$, and $(OH^-) = 10^{-9}$; the latter value is practically negligible. Since the solution as a whole must be electrically neu-

¹ The removal of small diffusible ions may be achieved by dialysis of the protein solution, enclosed in a semipermeable membrane, generally made of collodion or cellophane, against distilled water which is changed several times to remove the ions which pass out. To remove traces of ions which remain after dialysis, electrodialysis may be employed; the solution of the protein, or other macromolecule, is placed in a central compartment, separated from two outer compartments by semipermeable membranes. Electrodes are placed in the liquid in the outer compartments, and electrolysis of the solution removes ions which can pass through the membranes from the central compartment, leaving the macromolecule inside, since it cannot pass through. Various precautions are required in the practical operation of this procedure, which we shall not attempt to discuss here.

The use of a mixed-bed ion exchange resin is in many respects preferable to electrodialysis as a method of removing diffusible ions. The macromolecule solution flows down through a bed composed of two exchange resins, the particles of which are intimately mixed; one of the resins exchanges H^+ ions for the cations present in the solution; the other exchanges OH^- ions for the anions in the solution. The two resins are in the form of small particles, which are thoroughly mixed together, so that regions of H^+ ion exchange and regions of OH^- ion exchange are closely adjacent everywhere. A remarkably complete removal of ions may be achieved by this method. To minimize pH fluctuations during the process, a preliminary exchange is carried out with another mixed-bed resin, in which an anion such as acetate and a cation such as ammonium ion pass out into the solution from the resin, in exchange for other ions originally present. The buffering action of the ammonium and acetate ions then protects the solution from any marked pH fluctuations when it subsequently flows through the other mixed-bed resin which then replaces the NH_4^+ and Ac^- ions by H^+ and OH^- ions. A detailed procedure has been described by H. M. Dintzis (Thesis, Harvard University, 1952).

tral, the protein must carry a negative net charge, which on the average is $-0.1 (= 10^{-5}/10^{-4})$ per molecule. (For practical purposes this is virtually indistinguishable from zero.) The changes of pH observed, when various salts were added to the albumin at various concentrations, are shown in Fig. 2.

It will be seen that in every case the pH increases with increasing salt concentration, but the effects differ greatly from one salt to another. All the salts shown in Fig. 2 were sodium salts; it is known from other studies, discussed briefly above, that sodium ions are scarcely if at all bound by

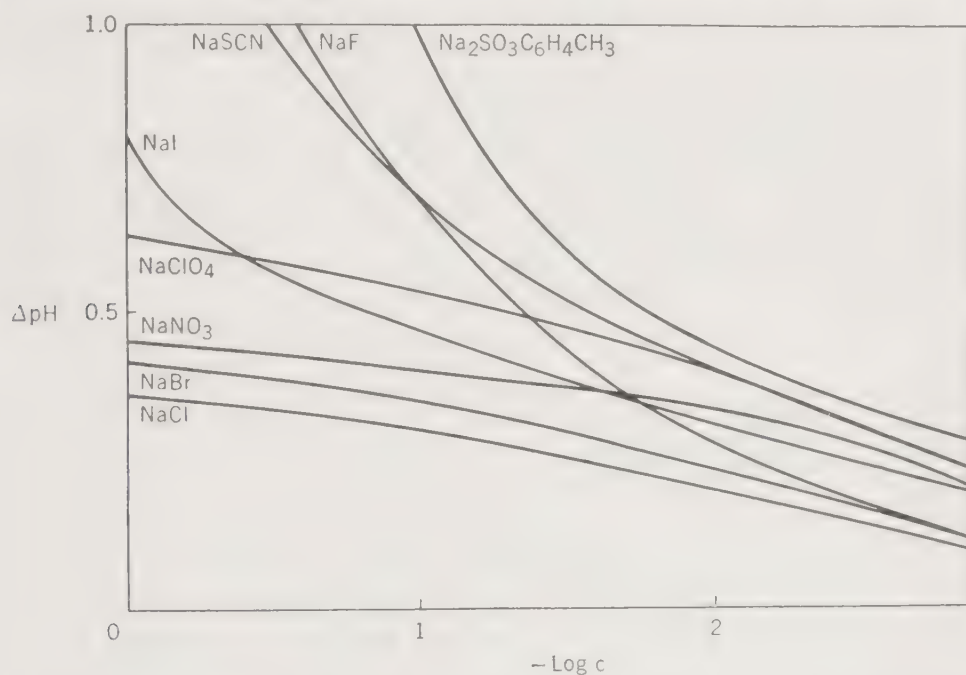


FIG. 2. Displacement of pH of isoionic serum albumin solutions by added salts. (From Scatchard and Black, 1949.)

albumin. The differences between the pH shifts for the various salts are therefore due to the differences in the binding of the anions. Those which bind strongly produce a large upward displacement of pH . Qualitatively the effect is simply explained. The binding of anions increases the attraction of the albumin for protons. More protons are therefore bound by the basic groups on the albumin, and since protons are thus removed from solution the pH rises. Since (H^+) is very small to begin with, and the albumin solution is too dilute to be a strong buffer, only a few protons need be bound to produce a large pH shift.

If we adopt the picture of the molecule as a uniformly charged sphere of radius b and net charge \bar{Z}_p , discussed in Chapter 9, then the net binding of ions may be calculated from the observed pH shift, from equation

(58) of that chapter. We consider the acidic groups in the albumin molecule to be divided into classes, the i th class of groups containing n_i groups, all characterized by a single intrinsic pk value $(pk_{\text{int}})_i$. Of these n_i groups, \bar{h}_i have lost protons at a given pH . Then equation (58) of Chapter 9 becomes

$$pH = (pk_{\text{int}})_i + \log \frac{\alpha_i}{1 - \alpha_i} - 0.868w\bar{Z}_p = (pk_{\text{int}})_i + \log \frac{\bar{h}_i}{n_i - \bar{h}_i} - 0.868w\bar{Z}_p \quad (9.58)$$

where w is defined by equation (53) of Chapter 9. A similar equation may be written for each of the classes of acidic groups in the molecule, each class having its own characteristic pk_{int} value. At a given pH , the sum of $(pk_{\text{int}})_i + \log \alpha_i/(1 - \alpha_i)$ must be the same for every class of groups, since w and \bar{Z}_p are parameters describing the molecule as a whole, and are the same for all groups. In the initially salt-free solution, as we have seen above, $\bar{Z}_p = 0$ (or, strictly speaking, $\bar{Z}_p = -0.1$, which we may regard as zero). Hence the initial pH , pH_0 , may be written

$$pH_0 = pk_{\text{int}} + \log \frac{1 - \alpha_i}{\alpha_i}$$

When salt is added, the mean net charge, \bar{Z}_p , becomes, if ions of the salt are bound by the protein,

$$\bar{Z}_p = \Sigma \bar{\nu}_k Z_k$$

Here $\bar{\nu}_k$ represents the mean number of ions of species k and valence Z_k bound by the protein. The change of pH on adding salt is always upward, for serum albumin solutions, and the pH in these experiments always lies between 5 and 6.5. Even if the pH changed from 5 to 7, the uptake of H^+ ions by the protein would be less than 0.07 per albumin molecule, in a solution containing 10 gram of albumin per liter. This figure is negligible compared to the number of ionizable groups in the molecule; there are for instance 17 imidazole groups in serum albumin, so we could set $n_i = 17$ for this class; and there are approximately 100 carboxyl groups (Chapter 9, Table VIII). With all these groups available, and the minute amount of H^+ ion which reacts, it is plain that α_i in equation (58) of Chapter 9 given above is not sensibly altered by the addition of salt. The shift of pH on adding salt must therefore be due entirely to the change of \bar{Z}_p from zero to $\Sigma \bar{\nu}_k Z_k$:

$$\Delta pH = pH - pH_0 = -0.868w\bar{Z}_p = -0.868w\Sigma \bar{\nu}_k Z_k \quad (4.1)$$

In the studies of Scatchard and Black it was known that the added cations were not bound, and the anions were univalent ($Z_k = -1$). Thus

we may simply write the mean number of anions bound as \bar{v} , and the above equation becomes

$$\Delta pH = 0.868w\bar{v} \quad (4.2)$$

To evaluate w , Scatchard *et al.* (1957) assumed that the albumin molecule could be described by a spherical model of radius $b = 30$ Å, and that the collision diameter, a , for albumin and the surrounding small ions was 32.5 Å. From equation (53) of Chapter 9 this gives for w , as a function of the ionic strength, ω , in water at 25°,

$$0.868w = 0.1034 - \frac{1.017 \sqrt{\omega}}{1 + 10.663 \sqrt{\omega}} \quad (4.3)$$

It is obvious from Fig. 2 and equation (4.2) that binding of anions to serum albumin increases in the order chloride < bromide < nitrate < iodide < thiocyanate < *p*-toluenesulfonate. Perchlorate and fluoride are also strongly bound, but they give curves which cross some of the other curves. In general the order of increasing binding corresponds to the Hofmeister series, which has been briefly discussed in Chapter 5, in connection with the salting-out effect.

Isoelectric and Isoionic Solutions

In Chapter 9 we discussed the concept of the isoelectric point but confined the discussion to amphoteric substances which were supposed to acquire a net charge only through binding or release of protons. On this assumption the isoelectric point, defined as the *pH* at which the mobility of the amphoteric substance (P) is zero, may be calculated from equations (45) and (46) of Chapter 9, if the ionization constants are known. In general we must reckon with the binding of ions other than protons, and the isoelectric point will depend on the concentrations of all the ions present in the solution, if they have any tendency to bind to P. The substance P in solution may exist in many different forms, depending on the number of protons attached to the various basic groups present, and the number of other ions which may be bound to P in any way whatever. Any individual species of P, the concentration of which we may denote by (P_i) , carries a net charge, Z_i , in proton units. The mean net charge of P is given by summing up the charges contributed by all species and dividing by the total concentration of all species:

$$\bar{Z}_p = \frac{\sum Z_i(P_i)}{\sum (P_i)} = \sum \frac{Z_i(P_i)}{m_p} \quad (4.4)$$

Here m_p is the total molar concentration of P. Then the condition for the isoelectric point is that $\bar{Z}_p = 0$. We might try to simplify our system by

removing all small ions, except of course H^+ and OH^- ions, by electro-dialysis or by treatment with a mixed-bed ion exchange resin. It is apparent from (4.4), however, that P in such a system cannot be isoelectric, except in the very special case that the isoelectric point coincides with the pH of neutrality. To make P isoelectric, there must be some anions present if the isoelectric solution is acid, or cations present if the isoelectric solution is alkaline, to balance the difference in concentration between H^+ and OH^- ions.

In discussing such systems it is convenient to define what is known as an isoionic solution. The simplest example of such a solution is one of the sort just referred to, which contained no ions except those of P, and H^+ and OH^- ions; this is exemplified by the solutions studied by Scatchard and Black, which we have just been discussing, before salt had been added to them. The condition for electrical neutrality in such an isoionic solution is

$$\sum Z_i(P_i) = (OH^-) - (H^+) = \bar{Z}_p m_p \quad (4.5)$$

where the second equality results from (4.4). It is obvious from (4.5) that, as m_p approaches zero, the isoionic pH must approach that of neutrality. We have considered one simple case of this sort in calculating the pH of aqueous solutions of glycine (Chapter 9, p. 507). In that case we saw that the isoionic pH became virtually identical with the isoelectric pH at all but very low concentrations of glycine. This is generally true for more complicated amphoteric substances also, if the isoelectric pH is not too far from neutrality. We have already calculated for $10^{-4} m$ serum albumin at pH 5, or slightly above, that $\bar{Z} = -0.1$, or less; this is immediately obvious from (4.5). A net charge of -0.1 is usually indistinguishable from zero experimentally; and in this case there is practically no difference between the isoelectric and the isoionic pH . If, however, the isoionic pH is 4, and m_p is still 10^{-4} , then $\bar{Z} = -1$, and the isoionic pH may be significantly different from (and higher than) the isoelectric pH . The discrepancy is of course even more marked if the isoelectric point is more acid than 4, and it becomes greater as m_p decreases, from (4.4) and (4.5). Similar reasoning holds for solutions alkaline to neutrality; here of course \bar{Z} is positive in the isoionic solution, and the isoionic solution is always at a lower pH than the isoelectric solution.

The concept of an isoionic solution may now be given a further extension; if a neutral salt is added to the salt-free solution of P which we have discussed above, the solution is still isoionic *by definition*. It ceases to be isoionic if an acid or base is added to the system. Thus if a careful analysis shows that the *total* concentration (bound and unbound) of all the cations in the system—excluding ions derived from P, H^+ , and OH^- ions, and other acids and bases—is equivalent to the *total* concentration of

all the anions, then the solution is isoionic. It is clear by this definition that the serum albumin solutions shown in Fig. 2 are isoionic solutions, and the definition of the isoionic point, as we have given it here, follows closely that of Scatchard and Black. It is also clear from Fig. 2, and from the electrophoretic data of Longworth on serum albumin mentioned above, that addition of neutral salt may shift the isoelectric and the isoionic pH values in opposite directions. Thus the addition of $0.1\ M$ sodium thiocyanate to salt-free human serum albumin is (see Fig. 2) accompanied by an upward shift of pH of approximately 0.75 (see equation 4.2). On the other hand, Longworth found the isoelectric point of bovine serum albumin in $0.01\ M$ acetate + $0.09\ M$ thiocyanate to be 4.17 , approximately 0.8 to $0.9\ pH$ unit lower than the isoelectric (or isoionic) pH in the absence of salt. (Here we are comparing measurements of one kind on human serum albumin with measurements of another kind on bovine albumin. The two proteins, however, although not identical, are so closely alike that the differences between them can be ignored for the purposes of the discussion here.) The reason for the upward shift of the isoionic pH with added salt has already been discussed. The downward shift of the isoelectric point can also be readily explained qualitatively as we have indicated above in the discussion of method 6.

The concept of isoionic solutions was first developed by Sørensen *et al.* (1927), who proposed two somewhat different definitions, one of which was essentially equivalent to the definition adopted here.

Some Fundamental Considerations Regarding Certain Types of Binding

We consider a molecule or ion (P), which contains n binding sites each of which is capable of attaching another molecule or ion, known as the ligand (A). It is not necessary that P be a large molecule; it may even be a simple monatomic ion, such as Fe^{++} , Fe^{+++} , Cu^{++} , or Zn^{++} , since all these are capable of binding several (four to six) water molecules, or basic molecules such as ammonia or imidazole. In a solution of hemoglobin and oxygen, for instance, P is hemoglobin and A is oxygen. In a solution of cupric ion and ammonia, P is the Cu^{++} ion, and A is ammonia. The sites on P , which may be occupied by A , are not necessarily empty when A is absent. Indeed for the Cu^{++} ion, and for most similar simple cations, it is known that in aqueous solutions water molecules are bound to the same sites to which ammonia can bind. There is evidence, from work of F. Haurowitz (1951), that in oxygen-free hemoglobin solutions a water molecule occupies the site on the iron atom which can also be occupied by oxygen. Thus these reactions involve a displacement of water by A from the binding sites, not a simple addition of A . Since, however, the amount of bound water is determined by the activity of water in the system (relative to the activity of A), and since the activity of water is virtually constant in dilute aqueous solutions, we can implicitly include the water activity in the measured association constants for the reaction between P and A , just as we have done for acid-base equilibria in Chapter 8. As the concentration of (A) in the solution is increased, the average number of sites on P which are occupied by A

must also increase, varying from zero when (A) is zero to a limiting value of n when (A) becomes infinite.

This average value has already been denoted in the discussion above by the symbol \bar{v} . It is denoted as an average, since in general, even if all the binding sites are equivalent, there will be a statistical distribution of the bound A molecules among all the available sites. For any given value of \bar{v} determined by experiment, there will at any moment be some P molecules which contain more than \bar{v} molecules of A bound, others less. The ratio \bar{v}/n gives the fraction of all the available sites which are occupied by molecules of A at any moment.

To have a reasonably adequate understanding of the situation, we should know (1) how many groups can be bound; that is, the value of n ; (2) how strongly they are bound; that is, the value of the affinity constant, k ; and whether a single constant, k , is sufficient to characterize all the n binding sites; (3) the nature of the various combining sites; (4) their locations in the structure of P; and (5) the biological significance of it all. That is, we wish to answer the questions put by Scatchard (1949) and by Scatchard *et al.* (1954): "How many? How tightly? Where? Why? What of it?"

Interactions of Bound Groups

In many cases we must ask other questions too. For instance, we may find that the binding of one molecule of A to P either helps or hinders the binding of another A molecule at a neighboring site; in the former case we speak of positive, in the latter of negative, interactions. The latter are the more commonly encountered; for example, all interactions observed in acid-base equilibria are negative, for the binding of one proton alters the electric charge on the molecule to which it is bound so that it requires more electrostatic work to put on another proton at a neighboring spot. In hemoglobin, on the other hand, placing an oxygen molecule on one heme group increases the tendency for its neighbors to pick up more oxygen. The same thing is commonly true for oxidations and reductions, which in organic molecules commonly involve an uptake of two electrons to go from the completely oxidized (T) to the completely reduced form (R). If only one electron is added to T, the intermediate form (S) is a free radical, known generally as a semiquinone. Semiquinones are often highly unstable; so that if S is formed from T by the addition of one electron, it then displays a much higher electron affinity than T did, and immediately picks up a second electron to form R. In many systems, indeed, S is practically undetectable; the system goes from T to R by what is virtually a simultaneous transfer of two electrons.

Competition Effects

In addition to P and A, the system may contain another molecule, B, which binds at the same sites as A. In that case, the amount of A bound, at a given activity of free A in solution, is also dependent on the activity of B. Hemoglobin again furnishes an excellent example. Oxygen and carbon monoxide compete for the same four heme groups; if both gases are present, the amount of either one bound is a function of the activity (or partial pressure) of the other, as well as of its own activity. The binding of metallic cations to proteins is another constantly encountered example; such ions compete with protons for basic sites on the protein molecule. If the *pH* goes up—that is, the H^+ ion activity goes down—at a given metal ion concentration, the binding of the latter ion is favored. The effect of competitive inhibitors in enzyme systems, which occupy the same active spot on an enzyme molecule at which the substrate is normally attached, is probably the outstanding example of competitive interactions in biochemistry.

When two kinds of molecules are competing for the same sites, it is generally true that increasing the concentration (or activity) of one decreases the amount of the other that is bound. This is not necessarily true, however; if positive interactions also exist, then the binding of one may actually facilitate the binding of the other. The hemoglobin-oxygen-carbon monoxide system is the prime example. If one heme group is occupied by either O_2 or CO, its neighbors in the same hemoglobin molecule have a greater affinity for either O_2 or CO than if the heme is not occupied by either kind of molecule. Thus at low oxygen pressures, where most of the heme groups are unoccupied, the introduction of carbon monoxide into the system at a very low pressure actually increases the amount of oxygen bound. At higher gas pressures, however, at which nearly all the heme groups are occupied by either O_2 or CO, the more usual type of competitive effect is observed, and raising the partial pressure of either gas tends to drive off some of the other from its attachment to the hemoglobin. This system is considered later in more detail (Volume II). So far as we are aware, other analogous systems have not yet been discovered in biochemistry, but the possibility of their existence should be borne in mind. Such systems might be well adapted for performing certain biological functions.

Interactions between Different Kinds of Bound Groups

Two different kinds of ligand molecules may be present—we may again call them A and B—which bind to P, but at different sites, so that there is no direct competition. Binding of B at one of the sites at which

it can be attached, however, may transmit an effect through the molecule of P, or through the surrounding medium, which alters the affinity for A of one of the sites on P which are specific for it. Such an effect is necessarily reciprocal; if binding of B increases the affinity of P for A, at constant (A) activity, the binding of A must increase the affinity of P for B at constant (B) activity. The mathematical relationships are given later (p. 654). Hemoglobin is again an outstanding instance; binding of oxygen to one of the heme groups increases the tendency of protons to dissociate, at constant pH , from one of the neighboring groups, generally known as a heme-linked acid group.² The reciprocal relation requires that the oxygen affinity of hemoglobin is increased when the heme-linked group loses a proton, at constant oxygen pressure. It should be noted that the heme-linked acid groups are a small and very special class; the great majority of the acidic groups in hemoglobin—nearly 200 in all—are apparently quite unaffected by the oxygenation of the hemes.

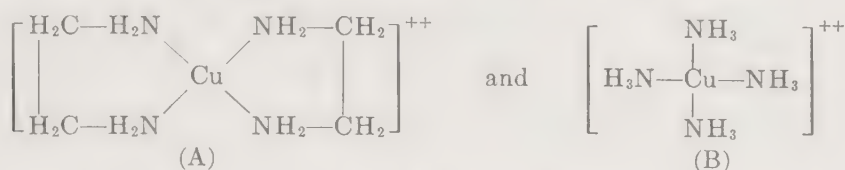
In serum albumin and other protein solutions, the simultaneous binding of protons and of anions furnishes another example. The nature of the loci at which anions, such as chloride or thiocyanate, are bound is uncertain; probably they are positively charged groups, such as ammonium, imidazolium, or guanidinium groups in certain specific configurations which favor anion binding. In any case, however, the binding of the anions makes the net charge on the protein more negative (or less positive); hence it favors the binding of protons, more of which will be taken up at a given pH if some anions are also bound. Here it is also obvious that the effect is reciprocal; decreasing the pH increases the binding of anions for electrostatic reasons. This effect, of course, is perfectly general, unlike the highly specific interaction of the heme and the heme-linked acid group in hemoglobin. It is in a way the converse of the competitive interaction, of which we have spoken, between protons and other cations for basic groups on proteins. Bound anions, however, are not competing for the same basic groups to which the protons attach themselves; instead the binding of protons to certain basic groups in the protein molecule—for instance, amino groups—is probably a pre-requisite for the binding of the anions.

Formation of Chelate Complexes

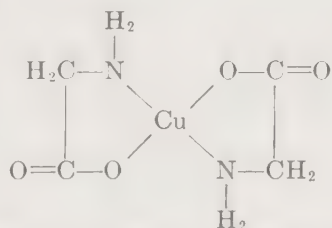
The ligand A, which attaches itself to P, may contain two, three, or more loci capable of forming an attachment. Such complexes are known

²More exactly, it appears that there are two such linked acid groups for each heme, one of which becomes more acidic, the other less acidic, on oxygenation of the heme. Under physiological conditions, however, it is the latter which is important (see the discussion of hemoglobin in Volume II).

as chelate compounds, and a vast number of them are now known (see, for instance, Martell and Calvin, 1952; Schwarzenbach, 1954). Generally they are much more stable than the complexes formed by simple ligands containing only one such reactive group. Ethylenediamine, for example, which can form a five-membered ring with a metallic ion such as Co^{++} , Cu^{++} , or Zn^{++} , gives complexes which are much more stable than those of ammonia with the same metals. Thus of the two structures:

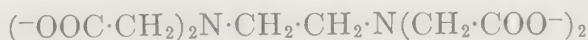


(A) is formed at concentrations of ethylenediamine much lower than the ammonia concentrations that are necessary to form (B). Much experience has shown that chelate complexes involving five-membered rings are particularly stable. Five-membered ring complexes are also formed by the anions of simple amino acids such as glycine, the nitrogen of the amino group and one of the oxygens of the ionized carboxyl forming the linkages to the metallic ion.



The Copper-Glycinate Complex

Even ions such as Ca^{++} , with considerably less tendency to bind to simple ligands than Co^{++} , Cu^{++} , or Zn^{++} , can form strong chelate complexes with "multidentate" ligands, of which ethylenediaminetetraacetate (EDTA, Versene), developed by Schwarzenbach, is the best known:



Here there are six groups—two nitrogen and four carboxylate groups—which are capable of coordinating with the calcium ion, which is practically enclosed in a cage when it forms part of the complex. Less powerful as a complexing agent for calcium, but still very effective, is the citrate ion. Because of this property citrate has found wide use in preventing the clotting of blood, by combining with the calcium ion which is essential for the conversion of prothrombin to thrombin.

The outstanding examples of chelate complexes in biochemistry, however, are undoubtedly the metalloporphyrins, especially chlorophyll, vi-

tamin B₁₂ (a cobalt porphyrin derivative),* and the heme derivatives, including the cytochromes, catalases, peroxidases, erythrocrucorins, and hemoglobins. Here the metallic atom sits in the middle of a square of nitrogen atoms, one from each of four pyrrole rings, these rings being linked together in turn by —C(H)= (methine) bonds into a single great ring. The whole structure lies in a single plane, except for the side chains attached to the outer carbon atoms of the pyrrole rings. This form of linkage for the central metallic atom, surrounded by a rigid structure with four coplanar coordinating nitrogen atoms arranged in a sequence at a suitable distance from the metal, is a particularly stable one. At the same time an atom such as iron, which normally forms six coordinate bonds directed toward the corners of a regular octahedron, still has two coordination positions free—one above the great ring, one below—for reaction with other compounds. In the heme proteins, one of these must be used to coordinate the heme to a group in the protein molecule; the other may also be so coordinated, but may instead be available, as in the hemoglobins, for reaction with other compounds such as oxygen. Thus the structure provides a high degree of stability for the metalloporphyrin complex, combined with the possibility of a high degree of reactivity with other compounds. It seems highly probable that other classes of metalloproteins and metalloenzymes in which the metal is very firmly bound to the protein are also chelate compounds. The nature of the binding of copper in the hemocyanins, for instance, is still unknown, as is that of zinc in carbonic anhydrase, carboxypeptidase, and alcohol dehydrogenase; but it may be predicted with considerable confidence that they will be found to be chelate complexes when their structure is unraveled.

Binding by a Set of Equivalent and Independent Groups

We now return to a more detailed analysis of binding by a molecule, P, containing n groups capable of attaching a ligand, A. We deal now with the simpler cases in which we may set the activities of the components equal to their concentrations, and especially with the idealized case in which all the n combining sites on P may be considered as equivalent and independent. First consider the special case when $n = 1$, so that $\bar{\nu}$, the average number of moles of A bound by one molecule of P, varies only between 0 and 1. Then the only reaction taking place is



* This statement is not precise. Two of the four pyrrole rings surrounding the cobalt are linked to one another directly, not through a methine bridge, so that this is not a true porphyrin structure, although closely related to it. The brilliant X-ray diffraction studies of D. Crowfoot Hodgkin and her associates have elucidated the structure in great detail. For a recent review see Kendrew, J. C., and Perutz, M. F. (1957). *Annual Rev. Biochem.* **26**, 327.

and the association constant, k_{assoc} , is given by

$$k_{\text{assoc}} = \frac{(\text{PA})}{(\text{P})(\text{A})} = \frac{1}{K_{\text{diss}}} \quad (5)$$

If A is a proton, this is of course the equation for the ionization of a monobasic acid, which has been discussed at length in Chapter 8. There and in Chapter 9 we have formulated the relations involved in terms of the dissociation constant, K_{diss} . Here we shall use chiefly its reciprocal, the association constant. For acid-base equilibria, setting $\text{A} = \text{H}^+$, Brönsted has termed k_{assoc} the "basicity constant"³ of the base P which is conjugate to the acid HP. We may note the numerical identity between the logarithms of these constants:

$$\log k_{\text{assoc}} \equiv pK_{\text{diss}} \equiv -\log K_{\text{diss}} \quad (6)$$

It is merely a matter of convention and convenience whether we use k_{assoc} or K_{diss} .

Considering all the molecules of P and PA in solution, we denote the fraction of binding sites on P, which are unoccupied, by the symbol α , as in the special case of acid-base equilibria (Chapter 8, equations 18 and 19):

$$\alpha = \frac{(\text{P})}{(\text{PA}) + (\text{P})} = \frac{K_{\text{diss}}}{(\text{A}) + K_{\text{diss}}} = \frac{1}{1 + k_{\text{assoc}}(\text{A})} \quad (7)$$

The average number of moles of A bound, for all the molecules of P and PA present, is $\bar{\nu} = 1 - \alpha$.

$$\bar{\nu} = 1 - \alpha = \frac{(\text{PA})}{(\text{PA}) + (\text{P})} = \frac{(\text{A})}{(\text{A}) + K_{\text{diss}}} = \frac{k_{\text{assoc}}(\text{A})}{1 + k_{\text{assoc}}(\text{A})} \quad (8)$$

Here $\bar{\nu}$ may also be regarded as the probability that any molecule of P, chosen at random from the solution, will be found to have a molecule of A attached to it. From (7) and (8) (compare Chapter 8, equation 18):

$$\begin{aligned} \log k_{\text{assoc}} &= -\log (\text{A}) - \log \frac{\alpha}{1 - \alpha} \equiv -\log (\text{A}) + \log \frac{\bar{\nu}}{1 - \bar{\nu}} \\ &\equiv p\text{A} + \log \frac{\bar{\nu}}{1 - \bar{\nu}} \quad (9) \end{aligned}$$

If P, instead of containing one site capable of binding A, contains n such sites which are equivalent and independent, the extension of these equations is very simple. If we denote by PA_i a molecule of P which con-

³ This should not be confused with the "basic dissociation constant," $K_{\text{B}} = K_{\text{w}}/K_{\text{A}}$, which is defined in Chapter 8, equation (5).

tains one molecule of A attached at the i th binding site, then the intrinsic association and dissociation constants for the formation of PA_i from P and A are, exactly as in (5),

$$(k_{\text{assoc}})_i = \frac{(PA_i)}{(P)(A)} = \frac{1}{(K_{\text{diss}})_i} \quad (10)$$

Since by hypothesis the extent of binding at other sites on P does not affect the extent of binding at site i , equation (10) describes the extent of binding at site i , averaged over *all* the molecules of P in the system. Hence the value of $\bar{\nu}_i$ at site i , which must lie between zero and unity, is given by an equation of exactly the form of (8):

$$\bar{\nu}_i = 1 - \alpha_i = \frac{(k_{\text{assoc}})_i(A)}{1 + (k_{\text{assoc}})_i(A)} \quad (8.1)$$

By our hypothesis, however, all the n groups are not only independent but also equivalent. Hence the values of k_{assoc} and K_{diss} are the same for all the groups on any molecule of P, and we may drop the subscript i .

If $\bar{\nu} = \sum_{i=1}^n \bar{\nu}_i = n\bar{\nu}_i$ is the total average number of occupied sites per molecule of P, then the probability that any site, chosen at random from any molecule in the solution, is occupied by A becomes

$$1 - \alpha = \frac{\bar{\nu}}{n} = \frac{k_{\text{assoc}}(A)}{1 + k_{\text{assoc}}(A)} \quad (11)$$

The argument here is identical with that already given for a polybasic acid in Chapter 9 (see equations 27 and 28 of that chapter), and the conclusion is of course the same. We note that the quantity \bar{h} , defined in equations (8) and (38) of Chapter 9, and employed throughout that chapter, is equal to $n - \bar{\nu}$ in terms of the notation we are using here.

In the rest of the discussion in this chapter, relations will be expressed in terms of association constants only, and for simplicity of notation we shall write simply k , instead of k_{assoc} . Thus (11) becomes

$$\bar{\nu} = \frac{nk(A)}{1 + k(A)} \quad (12)$$

or

$$\frac{\bar{\nu}}{n - \bar{\nu}} = k(A) \quad (12.1)$$

In general, of course, the situation will turn out to be more complex than this. Most macromolecules are likely to contain, not one single set

of binding sites, but several such sets, as in the case of the various classes of proton-binding sites found for the proteins discussed in Chapter 9. We may try to deal with the situation by assuming m different sets of sites, the first set with n_1 equivalent and independent groups, each with association constant k_1 , the second set with n_2 such groups, each with association constant k_2 , and so forth. We write the k values in order: $k_1 > k_2 > k_3 > \cdots > k_{m-1} > k_m$. The total number of binding sites is n :

$$n = n_1 + n_2 + \cdots + n_m = \Sigma n_i$$

and the mean number of sites occupied by molecules of A is

$$\bar{v} = \sum_{i=1}^m \frac{n_i k_i (A)}{1 + k_i (A)} \quad (13)$$

Equation (13), like (12), takes no account of interactions between the sites, which will be considered later.

If we divide both sides of (13) by (A) , and take the limit of $\bar{v}/(A)$ as (A) approaches zero, we obtain

$$\lim_{(A) \rightarrow 0} \frac{\bar{v}}{(A)} = \sum_{i=1}^m n_i k_i \quad (13.1)$$

On the other hand, if we suppose (A) in (13) to become infinite, it is obvious that \bar{v} approaches n as a limit:

$$\lim_{(A) \rightarrow \infty} \bar{v} = \sum_{i=1}^m n_i = n \quad (13.2)$$

Equation (13) includes endless possible sets of n 's and k 's, and it would be hopeless to attempt any general discussion of all these possibilities. We may, however, illustrate some of the important factors that arise, where there is more than one set of equivalent and independent groups, by considering one simple example.

We assume the macromolecules P to have $n_1 = 10$ binding sites with $k_1 = 1000$, and $n_2 = 30$ other sites with $k_2 = 25$. This corresponds to the values assumed by Scatchard *et al.* (1950) when P is human serum albumin and A is thiocyanate ion. In their calculations they made allowance for electrostatic interactions—we return to this problem on p. 645—but here we ignore such effects and deal with an idealized albumin-thiocyanate system in which electrostatic forces are absent. Then (13) becomes

$$\bar{v} = \frac{n_1 k_1 (A)}{1 + k_1 (A)} + \frac{n_2 k_2 (A)}{1 + k_2 (A)} = \frac{10,000(A)}{1 + 1000(A)} + \frac{750(A)}{1 + 25(A)} \quad (13.3)$$

On the other hand, if only the 10 sites with the highest k value were present, the second term on the right of (13.3) would drop out, and (13.3) would reduce to a special case of (12).

In general, when we are studying binding in a system not previously explored, we determine $\bar{\nu}$ and (A) by one of the methods already outlined, or sometimes by some other technique. The problem then is to infer the number of binding sites, n , and the affinity constant, k , if there is only one class of binding sites. In the more general case we have to determine several different constants, $k_1, k_2 \dots k_m$, and the number of sites corresponding to each. We now consider several different methods of analyzing the experimental data in order to evaluate the n 's and k 's. We consider in each method first the result to be expected when there is only one set of sites (equation 12), and then the modifications produced by the presence of a second set of more weakly binding sites (equation 13.3).

THE "DIRECT PLOT" OF $\bar{\nu}$ AS A FUNCTION OF FREE LIGAND CONCENTRATION

The most obvious way to handle the data is to plot $\bar{\nu}$ against (A) . The curve obtained from (12) for a single set of equivalent groups is a segment

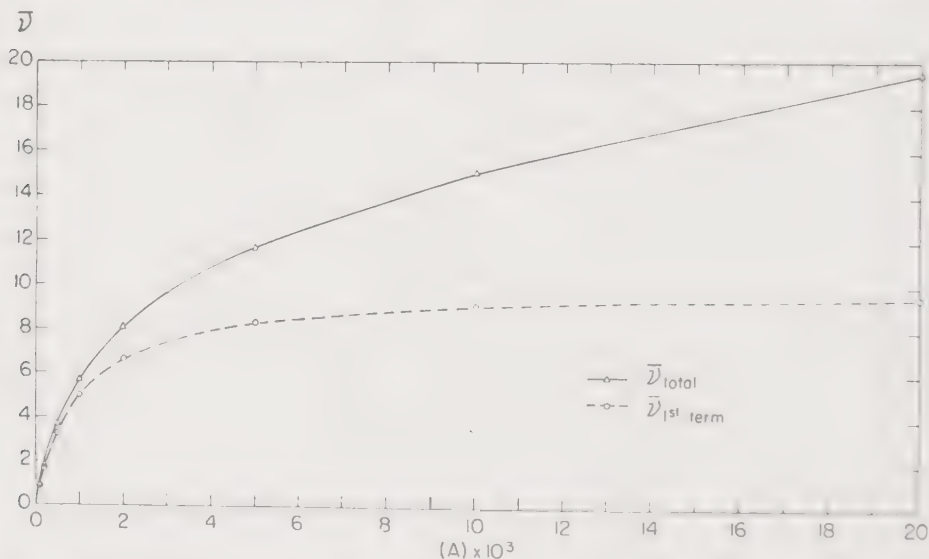


FIG. 3. Plot of $\bar{\nu}$ against (A) . Broken line for a single set of 10 groups with $k = 1000$ (equation 12). Solid line, for two sets of groups, as defined by equation (13.3).

of a rectangular hyperbola, which passes through the origin (see the lower curve of Fig. 3). The limiting slope is $\bar{\nu}/(A) = nk$ from (13.1); at high (A) values the curve levels off and approaches asymptotically the horizontal straight line $\bar{\nu} = n$. At $\bar{\nu} = n/2$ it is clear from (12) that $(A) = k^{-1}$ (or K_{diss}). This type of curve is often useful in representing certain data for

simple cases; for example, it represents well the binding of oxygen by myoglobin, the hemoglobin of muscle, which contains only one binding site for oxygen. In the lower curve of Fig. 3 we have taken $n = 10$, $k = 1000$; saturation of the binding sites is about 96% complete at the highest (A) value shown and is 50% when $(A) = k^{-1} = 0.001$. The added presence of a second set of 30 binding groups with $k_2 = 25$ (equation 13.3) leads to little change in the limiting slope, from 10,000 to 10,750 (see 13.1), but $\bar{\nu}$ continues to climb and, at $(A) = 0.02$, $\bar{\nu}$ is still less than half the saturation value of 40. Clearly, in most cases, this method of representing the data is not likely to lead to accurate values of the n 's or k 's.

THE LOGARITHMIC PLOT: BJERRUM'S "FORMATION FUNCTION"

Instead of plotting $\bar{\nu}$ against (A), we may plot it against $\log (A)$ or against $-\log (A) \equiv pA$. Jannik Bjerrum has called this method of representing the data a "formation function." For acid-base equilibria, with a monobasic acid, this is identical with the usual titration curve in which pH is plotted as abscissa, and bound protons as ordinate, taking $\bar{\nu} = 0$ when the substance being titrated is in its most basic form. From (12.1), taking logarithms, we may write for a single set of equivalent groups:

$$\log k + \log (A) = \log \frac{\bar{\nu}}{n - \bar{\nu}} \quad (14)$$

The curve for n equivalent and independent groups is shown, for $n = 10$ and $k = 1000$, in the lower curve of Fig. 4; it is obviously a simple titration curve (compare Chapter 9, Fig. 8, the curve for $w = 0$). The superimposed effect of the second set of 30 groups (equation 13.3) is shown in the upper curve of Fig. 4. The two curves coincide closely for values of $\log (A)$ below -3.5 but diverge progressively at higher values of $\log (A)$.

The representation of the data in Fig. 4 has an advantage over that of Fig. 3 in that a much wider range of values of (A) is conveniently represented. When examining experimental data, it is almost always worth while to plot them in this form; if the data approximate at all closely to the ideal type of curve shown by the broken line in Fig. 4, inspection of the curve allows at least an approximate estimate of the number, n , of binding sites. If n is known, the value of $\log (A)$ at which $\bar{\nu} = n/2$ immediately gives $-\log k$. If the experimental curve differs distinctly from the ideal curve, the deviations serve to indicate whether there are marked interactions between the groups, or whether the groups fall into two or more different classes, with different k values as in (13). The plot of the formation function, however, is often quite inadequate to determine n , if this was uncertain to begin with.

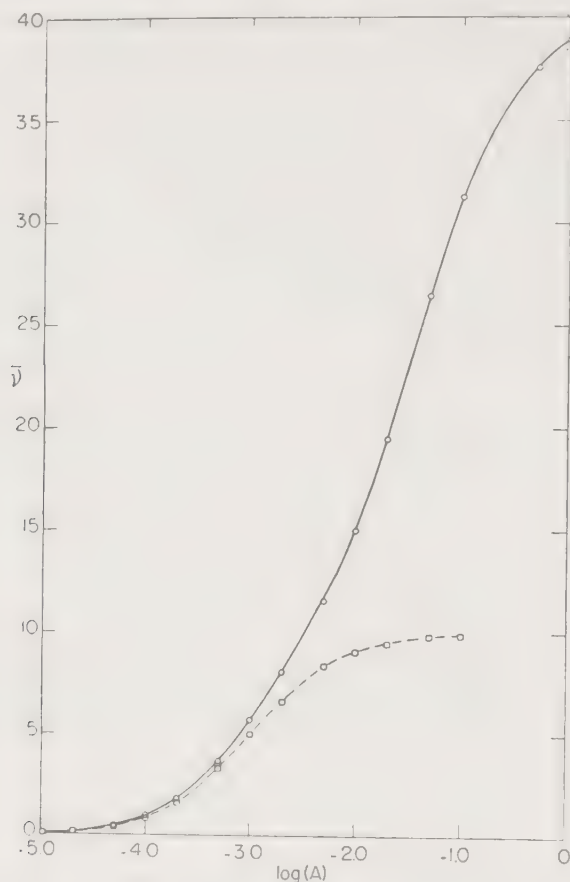


FIG. 4. Plot of $\bar{\nu}$ against $-\log(A)$; see equations (12) and (13.3). Broken line, for a single set of 10 groups with $k = 1000$. Solid line for two sets of groups. The data are the same as those of Fig. 3, but the range of (A) concentrations covered is of course far wider.

THE RECIPROCAL PLOT

If both sides of (12) are inverted we obtain

$$\frac{1}{\bar{\nu}} = \frac{1}{n} + \frac{1}{nk(A)} \quad (15)$$

A plot of $1/\bar{\nu}$ against $1/(A)$ gives a straight line if (15) holds, the intercept on the ordinate axis giving $1/n$; the slope is $1/kn$. An example, for $n = 10$ and $k = 1000$, is shown in the upper curve of Fig. 5. If the binding sites are indeed equivalent and independent, this plot determines both n and k directly. In more complicated cases, however, the method may give misleading results. Thus the lower curve of Fig. 5, which is for the two sets of binding sites described by equation (13.3), appears quite reasonably linear at $1/(A)$ values above about 500. Extrapolation of this

portion of the curve would cut the ordinate axis at $1/\bar{v} = 0.07$ to 0.08 approximately, giving $n = 13$ to 14 instead of the true value of 40 as $(A) \rightarrow \infty$ from (13.3) . It is only at very low values of $1/(A)$ —that is, at high (A) values—that a marked deviation from linearity occurs, and the curve bends downward to meet the ordinate axis at $1/\bar{v} = 0.025$. In dealing with actual experimental data, it is likely to be very difficult to tell whether this sort of deviation is occurring or not, at high (A) values.

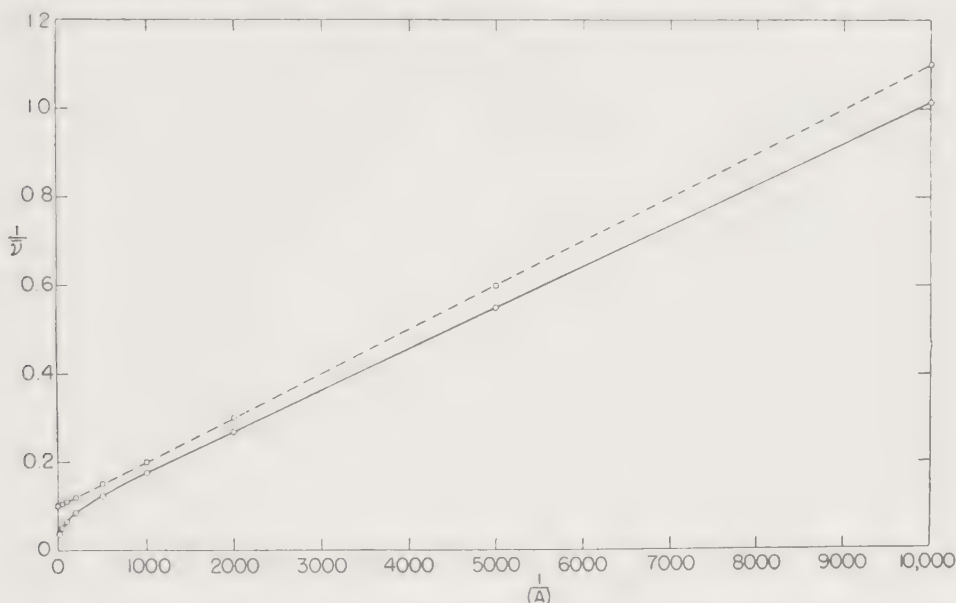


FIG. 5. The plot of $1/\bar{v}$ as a function of $1/(A)$; see equation (15). Same data as in Figs. 3 and 4.

THE SCATCHARD PLOT OF $\bar{v}/(A)$ AGAINST \bar{v}

Scatchard (1949) employed a different type of plot, which is also a straight line if the groups are equivalent and independent. If we multiply both sides of equation (12.1) by $(n - \bar{v})/(A)$, we obtain

$$\frac{\bar{v}}{(A)} = k(n - \bar{v}) \quad (16)$$

Then a plot of $\bar{v}/(A)$ as ordinate against \bar{v} as abscissa gives a straight line of the type shown in the lower curve of Fig. 6, in which the same data shown in the upper curve of Fig. 5 have been used. The intercept on the abscissa axis gives n , the maximum number of groups; the intercept on the ordinate axis gives kn . It should be noted that $\bar{v}/(A)$ approaches its maximum limiting value as (A) approaches zero, whereas at high (A) values, even though \bar{v} is increasing, $\bar{v}/(A)$ is approaching zero as \bar{v} approaches n . This method of analyzing the data lays less stress than the

reciprocal plot on the values of \bar{v} obtained at very low (A) values and gives more even relative weight to the different points on the curve. The plot of $\bar{v}/(A)$ against \bar{v} is also useful if there are several different classes of binding sites. As \bar{v} approaches zero, which occurs when (A) approaches zero, we see from (13.1) that $\bar{v}/(A)$ must approach an upper limit of $\Sigma n_i k_i$. On the other hand, when $\bar{v}/(A)$ approaches zero, which occurs as

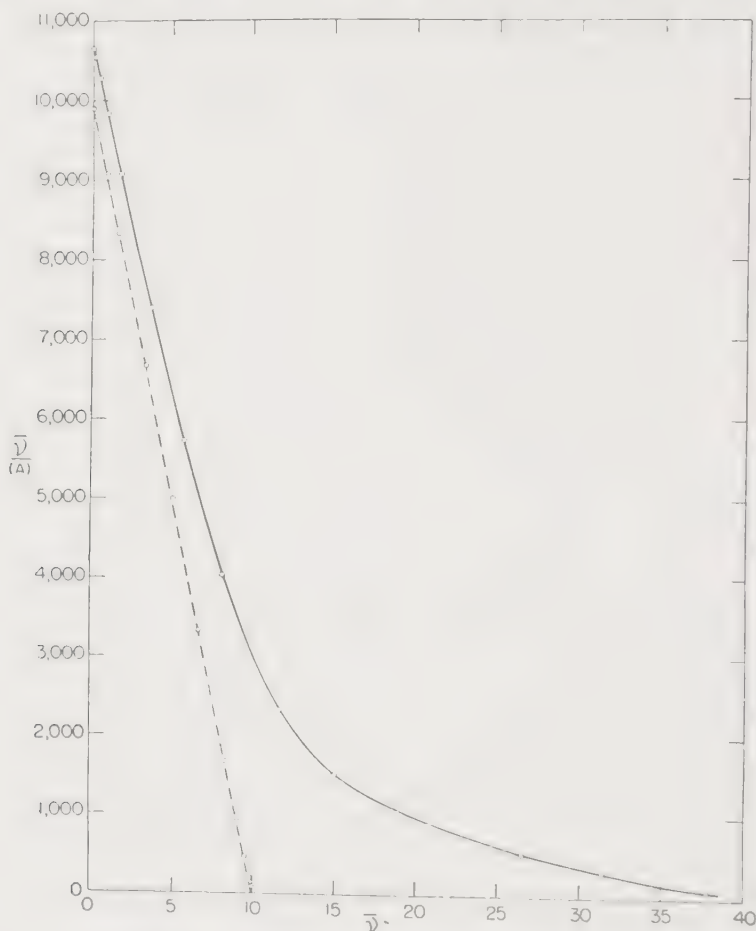
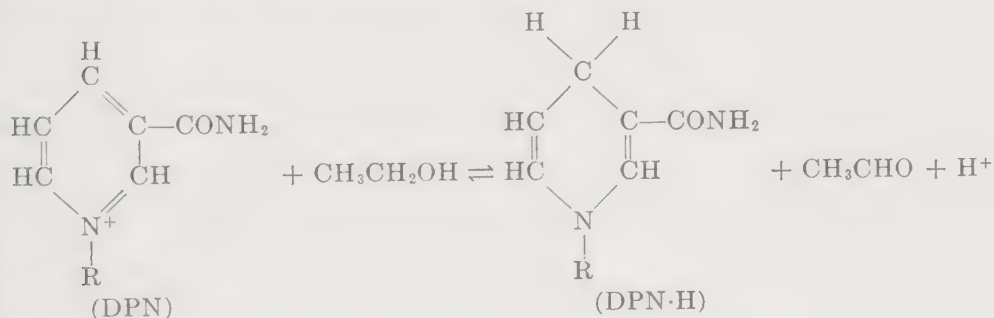


FIG. 6. The plot of $\bar{v}/(A)$ against \bar{v} ; see equation (16). Same data as in Figs. 3, 4, and 5.

(A) becomes very large, it is apparent from (13) that \bar{v} approaches a limiting value of $n = \Sigma n_i$. These relations are illustrated, for the system defined by equation (13.3), by the upper curve in Fig. 6. It is plain that both the abscissa and the ordinate intercepts are much more difficult to determine accurately for the markedly nonlinear curve, which is obtained in this case. In spite of the difficulty of the extrapolation, it is plainly far easier to obtain approximate estimates of the various values of n and k by this method than by the reciprocal plot, if more than one

set of binding groups is involved. A method of analyzing the experimental data, in cases where more than two sets of k values are needed to describe the results of experiment, has been presented by Scatchard *et al.* (1957), who have applied it to determine three different association constants for bovine serum albumin with chloride, thiocyanate, or other ions. Since this work involves electrostatic interactions, it is considered later in this chapter (p. 650).

We may further illustrate the analysis of binding data by a system of great biochemical interest, in which apparently only one set of equivalent groups is involved. Hayes and Velick (1954) have studied the combination of yeast alcohol dehydrogenase with reduced diphosphopyridine nucleotide (DPNH). The dehydrogenase is a protein of molecular weight of 150,000; the reaction which it catalyzes is, in one direction, the reduction of acetaldehyde to ethanol by DPNH; in the other direction, the oxidation of ethanol to acetaldehyde by DPN.



The large group denoted by R in the formula contains two residues of ribose, linked by a pyrophosphate group, and one residue of adenine at the far end of the molecule.

Even in the absence of the substrates acetaldehyde or ethanol, the enzyme binds reversibly both DPN and DPNH. The extent of binding was studied in the ultracentrifuge, the enzyme with attached coenzyme being separated by sedimentation from the free coenzyme in the surrounding liquid, as in the studies of calcium binding to casein which have already been discussed. The results were plotted in terms of Scatchard's equation (16).

Some of the data for binding of DPNH, in the presence and absence of DPN, are shown in Fig. 7. The extrapolated value of $\bar{\nu}$ (or \bar{r} in the notation of Hayes and Velick) is approximately 3.6. In other experiments the binding of DPN was also determined in the absence of DPNH; the scatter of the data was greater in this case, but the extrapolated limiting value of $\bar{\nu}$ was also compatible with the figure of 3.6. The association constant, k , for DPNH was evaluated from the abscissa and ordinate inter-

cepts by equation (16) as 7.7×10^4 . The value of k for binding of DPN was lower and considerably more uncertain; it was near 4×10^3 .

The experimentally found limiting value of \bar{v} (3.6) cannot be the true value of n ; the latter should be an integer if the enzyme preparation is homogeneous. Hayes and Velick concluded from their evidence that some of the enzyme preparation consisted of inactive protein, and that if this were allowed for the true value of n for the active enzyme was 4. This conclusion has been greatly strengthened by the later work of Vallee and Hoch (1955), which has shown that yeast alcohol dehydrogenase is a metalloenzyme containing four moles of zinc per mole of protein, and that

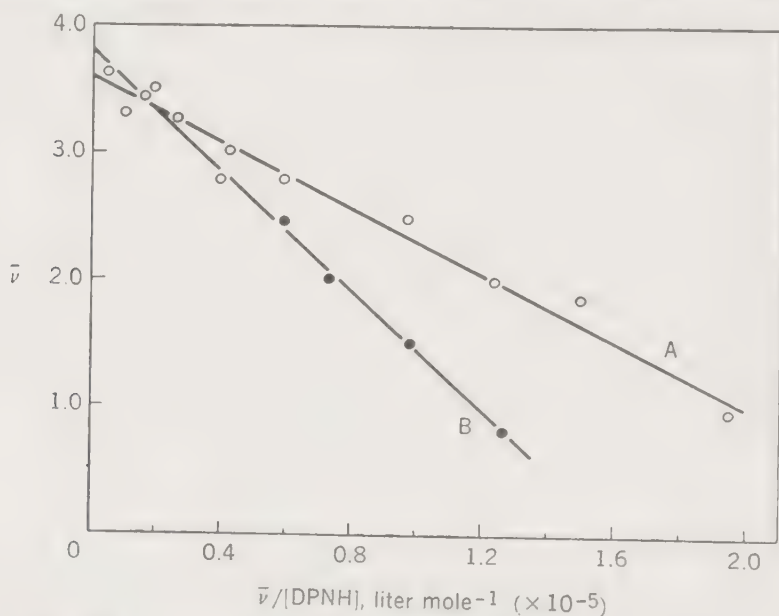


Fig. 7. Binding of DPNH by yeast alcohol dehydrogenase, plotted by the relation shown in Fig. 6. The lower straight line shows the effect of adding oxidized DPN, which competes with DPNH for the binding sites on the enzyme. (From Hayes and Velick, 1954.)

the activity of the enzyme is reversibly inhibited by reagents which combine with zinc, such as dithizone, 1,10-phenanthroline, 8-hydroxyquinoline, α, α' -dipyridyl, and thiourea. The zinc is bound very tightly to the protein, but the nature of the linkages involved is not yet established.

The work of Hayes and Velick also involved important kinetic studies of the action of this enzyme system; but these will not be considered here.

Relation of Equations for Binding to Those of Enzyme Kinetics

There is a close formal analogy—indeed in many respects an identity—between the equations just developed for reversible equilibria and those which have been developed to describe the kinetics of enzyme-substrate

interaction, beginning with the work of Michaelis and Menten in 1913. This is natural, for the formally identical elements of both are direct consequences of the law of mass action. It is worth while here to anticipate the discussion of enzyme kinetics in a later volume, and briefly state the relations involved. Consider an enzyme, E, which can combine with a substrate, S, to form a complex ES. The latter may dissociate reversibly into E and S, or it may undergo reaction to form the products of the reaction (which we denote by P), releasing E again. The velocity of formation of ES is proportional to the concentrations of E and of S, the factor of proportionality being the velocity constant, k_1 . The rate of decomposition of ES is proportional to (ES); the decomposition to form E and S is characterized by a velocity constant k_2 ; the decomposition leading to the formation of E and P by a velocity constant k_3 . Then the rate at which (ES) is changing with time is given by

$$\frac{d(\text{ES})}{dt} = k_1(\text{E})(\text{S}) - (k_2 + k_3)(\text{ES}) \quad (17)$$

The system attains a steady state, in which (ES) is not changing with time, soon after E and S are mixed, provided (as is usually the case) that $(\text{S}) \gg (\text{E})$, so that (S) is practically independent of time for a significant period after reaction begins. For the steady state, then, $d(\text{ES})/dt = 0$, and (17) becomes

$$\frac{(\text{E})(\text{S})}{(\text{ES})} = \frac{k_2 + k_3}{k_1} = K_M \quad (18)$$

where K_M is known as the Michaelis constant. It has the dimensions of concentration. The velocity, v , of the enzyme-catalyzed reaction by which the products, P, are formed is equal to $v = k_3(\text{ES})$. The total concentration, T , of enzyme in the system is equal to $(\text{E}) + (\text{ES})$. Given these relations, we may readily transform (18) into the equation

$$v = k_3(\text{ES}) = \frac{k_3 T (\text{S})}{(\text{S}) + K_M} \quad (19)$$

The constant K_M may be evaluated by determining v , the *initial* velocity of the enzyme-catalyzed reaction, as a function of the substrate concentration. At low (S), when $(\text{S}) \ll K_M$, v is proportional to (S). On the other hand, at very high (S), K_M in the denominator of (19) becomes negligible, and the velocity approaches a limiting maximum value, V_{\max} , equal to $k_3 T$. Under these conditions, every active center in the enzyme molecules is combined with substrate, and addition of further substrate can have no effect in accelerating the reaction. If we substitute V_{\max} for

k_3T in (19), it becomes

$$\frac{v}{V_{\max}} = \frac{(S)}{(S) + K_M} = \frac{K_M^{-1}(S)}{1 + K_M^{-1}(S)} \quad (20)$$

Here K_M^{-1} is simply the reciprocal of the Michaelis constant. Since K_M has the same dimensions as a dissociation constant, K_M^{-1} has the dimensions of an association constant. Thus (20) becomes formally identical with (12), if we replace \bar{v}/n by v/V_{\max} , (A) by (S), and k by K_M^{-1} .

Thus, just as (A) becomes equal to K^{-1} when $\bar{v} = n/2$, so (S) = K_M when $v = V_{\max}/2$. By taking the reciprocal of both sides of (20), and dividing by V_{\max} , we obtain

$$\frac{1}{v} = \frac{1}{V_{\max}} + \frac{K_M}{(S)V_{\max}} = \frac{1}{V_{\max}} + \frac{1}{(S)V_{\max}K_M^{-1}} \quad (21)$$

Thus, if $1/v$ is plotted against $1/(S)$, the curve is a straight line; the intercept, when $1/(S) = 0$, gives $1/V_{\max}$, and the slope is K_M/V_{\max} , or $1/V_{\max}K_M^{-1}$. This equation, which was derived by Burk and Lineweaver in 1934, has been widely used in expressing the results of enzyme kinetic studies. It is obviously an exact analog of (15), bearing the same relation to it that (20) does to (12). Likewise the kinetic data can be expressed in a form which is exactly analogous to Scatchard's equation (16):

$$\frac{v}{(S)} = \frac{V_{\max} - v}{K_M} = K_M^{-1}(V_{\max} - v) \quad (22)$$

Thus a plot of $v/(S)$ against v gives a straight line with an intercept $v = V_{\max}$ when $v/(S) = 0$, and an intercept V_{\max}/K_M , or $V_{\max}K_M^{-1}$, when $v = 0$. This method of plotting the results of enzyme kinetic studies was first employed by Eadie (1942).

The extension of these equations to deal with the action of competitive and noncompetitive inhibitors, and with other more complex systems, will not be given here. We note, however, that just as \bar{v}/n gives the average fraction of the binding sites on the molecule, P, which are occupied by the ligand, A, so $v/V_{\max} = (ES)/T$ gives the average fraction of the combining sites on the enzyme, E, which are occupied by the substrate, S.

As we have formulated them, the equations appear to imply that there is only one active site per enzyme molecule which can combine with substrate. They apply equally well, however, if each enzyme molecule contains n such sites, provided they are equivalent and independent. In that case, T in equation (19) is equal to n times the molar concentration of enzyme, and (A) and (ES) are n times the molar concentrations of free

and combined enzyme, respectively. Thus the value of n for yeast alcohol dehydrogenase is 4, as we have already seen. It is not at all sure, however, whether these 4 combining sites are truly equivalent and independent.

There is a fundamental difference, however, between K_M (or K_M^{-1}) and the dissociation (or association) constants occurring in equations such as (12), (15), and (16). The latter are true equilibrium constants. On the other hand the equilibrium constant for the reaction $ES \rightleftharpoons E + S$ is equal, in the notation of equations (17) and (18), to the ratio k_2/k_1 of two velocity constants. The measured value of K_M , however, is given from (18) by the ratio $(k_2 + k_3)/k_1$. This may become equal to the equilibrium constant defined above, but only if $k_3 \ll k_2$. This is sometimes found to be true in practice for certain enzyme systems, but in other cases it is found that $k_3 \gg k_2$, or that the two constants are of the same order of magnitude. The analysis of the enzyme system deals with a system in which a process is proceeding in a steady state. There is a fundamental distinction, of course, between such a system and one which is in true equilibrium. Nevertheless it is very useful to recognize the formal equivalence of the equations used to describe the two types of systems, since any advance in the theoretical analysis of one can be readily carried over to the other, provided that due account is taken of the physical significance of the mathematical symbols employed.

General Equation for Binding by a Molecule or Ion with n Sites Available for Combination

In general, if a molecule contains n reactive sites, at which combination with a ligand, A, can occur, the binding of A may be characterized by n association constants. The analysis of such a process has already been given in detail in Chapter 9, for the case in which A is H^+ ion. Here, however, we choose to formulate the relations in terms of association instead of dissociation constants. Thus, if the molecule, P, contains n binding sites for A, the association constants are

$$k_1 = \frac{(PA)}{(P)(A)}; k_2 = \frac{(PA_2)}{(PA)(A)} \cdots; k_i = \frac{(PA_i)}{(PA_{i-1})(A)} \cdots; k_n = \frac{(PA_n)}{(PA_{n-1})(A)} \quad (23)$$

Here (PA) denotes the sum of the concentrations of all the n microscopically different species in which one A molecule is bound to P at any one of the n sites, and (PA_i) denotes the corresponding sum for all the $n!/i!(n-i)!$ species in which i of the n sites are occupied by A. Thus k_1 , k_2 , etc., correspond, in the special case of acid-base equilibria, to the reciprocals of the macroscopic dissociation constants, K_1 , K_2 etc., defined

in Chapter 9, equations (31), (33), (34), and (38), if we set $k_1 = 1/K_n$, $k_2 = 1/K_{n-1}$, . . . , $k_n = 1/K_1$. The relation between these constants and the corresponding microscopic constants has already been discussed in detail in Chapter 9; the same analysis holds here and need not be repeated.

For compactness in notation, we also define the constants $L_0^* = 1$, $L_1^* = k_1$, $L_2^* = k_1 k_2$, . . . , $L_n^* = k_1 k_2 \cdots k_n$. Here again there is a reciprocal relation to the constants $L_1, L_2 \dots L_n$, defined in Chapter 9, equation (38)—namely, $L_1^* = 1/L_n$, $L_2^* = 1/L_{n-1}$, . . . , $L_n^* = 1/L_1$. Then the general expression for $\bar{\nu}$ is

$$\begin{aligned}\bar{\nu} &= \frac{(PA) + 2(PA_2) + \cdots + (n-1)(PA_{n-1}) + n(PA_n)}{(P) + (PA) + (PA_2) + \cdots + (PA_{n-1}) + (PA_n)} \\ &= \frac{k_1(A) + 2k_1 k_2(A)^2 + \cdots + nk_1 k_2 \cdots k_n(A)^n}{1 + k_1(A) + k_1 k_2(A)^2 + \cdots + k_1 k_2 \cdots k_n(A)^n} \\ &= \frac{\sum_{i=0}^n i L_i^*(A)^i}{\sum_{i=0}^n L_i^*(A)^i} = \frac{\partial \ln \sum_{i=0}^n L_i^*(A)^i}{\partial \ln (A)}\end{aligned}\quad (24)$$

If all the sites are equivalent and independent, so that the reaction at each site may be characterized by a single intrinsic microscopic constant, κ , which is the same for all, then by the same reasoning which leads to equations (39) and (40) of Chapter 9, we have

$$k_1 = n\kappa; k_2 = \frac{(n-1)\kappa}{2}; k_3 = \frac{(n-2)\kappa}{3}; \cdots; k_n = \frac{\kappa}{n} \quad (25)$$

In this case (24) reduces to the simple equation (12). The factors n , $(n-1)/2$, $(n-2)/3$, etc., in (25) are statistical factors. For example, k_1 is n times as great as κ in the limiting case given above, because P contains n acceptor sites for A , whereas any one of the microscopic species of the class PA contains only one donor site from which A can be released. Similarly PA contains $n-1$ receptor sites for A , but PA_2 contains two donor sites from which A can be released; the ratio of these two factors is $(n-1)/2$. These statistical factors are of course always present in any equilibrium of this type, whether or not the groups are equivalent and independent. It is therefore often illuminating to express the binding constants so that they are corrected for these statistical factors. We shall denote the resulting constants by the symbols $\kappa_1, \kappa_2 \dots \kappa_n$.

$$\kappa_1 = \frac{k_1}{n}; \kappa_2 = \frac{2k_2}{n-1}; \cdots; \kappa_{n-1} = \frac{(n-1)k_{n-1}}{2}; \kappa_n = nk_n \quad (26)$$

Since, for equivalent and independent groups, we should have $\kappa_1 = \kappa_2 = \dots = \kappa_i = \dots = \kappa_n$, observed differences between the successive κ values, determined on any actual system, indicate either intrinsic differences between the affinities of the individual groups, or interactions between the groups, or both. We now turn to a consideration of such systems.

The same reasoning which leads to equation (39.1) of Chapter 9 gives a relation between the association constants, k_1, k_2, \dots, k_n , or $L_1^*, L_2^*, \dots, L_n^*$, and a set of "titration constants," g_1, g_2, \dots, g_n , for the association of P with n molecules of A:

$$\sum_{i=0}^n L_i^*(A)^i = [1 + g_1(A)][1 + g_2(A)] \dots [1 + g_n(A)] \quad (27)$$

If A is H^+ ion, the g 's are related to the titration constants for dissociation, G_1, G_2, \dots, G_n , as defined in Chapter 9, equation (39.1), by the equations: $g_1 = 1/G_n, g_2 = 1/G_{n-1}, \dots, g_n = 1/G_1$. If we expand the product on the right-hand side of (27) and equate the coefficients of corresponding powers of (A) on both sides of the equation, the relations between the successive L^* (or k) values and the corresponding g values are immediately given. The titration constants g_1, g_2, \dots, g_n are real positive numbers, however, only if they form a progressive series in descending order of magnitude, or—in the limiting case of equivalent and independent groups—are all equal ($g_1 \geq g_2 \geq \dots \geq g_n$). This is always true for an acid-base titration, except for the rare cases in which the binding (or release) of one proton leads to a molecular rearrangement resulting in the formation of a new acidic group of greater strength than that which existed in the original molecule, P, at the start of the titration. It is probably true also for many other systems.

There are systems, however, which show positive interactions between the successive binding constants. Some such systems are considered later in this chapter, others in the next volume. For these the observed binding curves for \bar{v} as a function of (A) cannot be reproduced by any set of real positive values for g_1, g_2 , etc. Without attempting to prove this proposition in general, we may illustrate it for the simple case of a molecule, P, containing two binding sites. For $n = 2$, equation (27) becomes

$$\sum_{i=0}^2 L_i^*(A)^i = 1 + k_1(A) + k_1k_2(A)^2 = 1 + (g_1 + g_2)(A) + g_1g_2(A)^2 \quad (28)$$

Hence $k_1 = g_1 + g_2$, and $k_1k_2 = g_1g_2$ (compare Chapter 9, equation 18). In the limiting case where the groups are equivalent and independent, $g_1 = g_2$, and $k_2 = k_1/4$ (see Chapter 9, equations 21 and 22). In other cases we may write $k_2 = \beta k_1/4$, where the factor β takes account of the inherent differences between the groups and of the interaction between them. Hence $k_1k_2 = \beta k_1^2/4 = g_1g_2$, and $g_2 = \beta k_1^2/4g_1$. This permits us to express g_1 as a function of k_1 and β :

$$k_1 = g_1 + g_2 = g_1 + \frac{\beta k_1^2}{4g_1} \quad (29)$$

or

$$g_1^2 - k_1g_1 + \frac{\beta k_1^2}{4} = 0$$

An expression of identical form is obtained as the solution for g_2 . This quadratic, when solved for g_1 or g_2 , gives

$$g_1 \text{ or } g_2 = \frac{k_1(1 \pm \sqrt{1 - \beta})}{2} \quad (30)$$

Since $g_1 \geq g_2$, the plus sign in (30) corresponds to g_1 , the minus sign to g_2 .

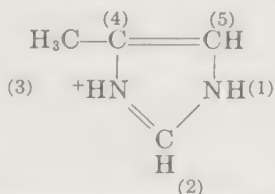
Obviously the value of g_1 can be real only if $\beta \leq 1$. For two equivalent groups with positive interactions, however, β must be greater than unity, and g would be imaginary. If the two groups were intrinsically different in their binding constants, β might be less than unity, even if there were positive interactions between the groups.

The physical meaning of these mathematical relations should be clearly understood. The titration constants g_1, g_2, \dots, g_n are the association constants which would be obtained for a system in which the actual molecule, P, with its n combining groups, is replaced by n hypothetical molecules, each with one combining group. Each of these groups is assigned an association constant g_1, g_2 , etc., in such a way that the binding of A due to all these molecules, in solution together, fits the observed curve for the binding of A to P. Two or more equivalent groups with negative interactions give a binding curve, for $\bar{\nu}$ as a function of $\log(A)$, which can be matched by a curve for two independent groups with different g values. The binding curve which rises most steeply at the mid-point, as $\log(A)$ increases, for independent groups, is obtained when the two groups are equivalent. A curve which rises more steeply than this can be obtained only if the binding groups interact positively—that is, in such a way that the binding of one A molecule promotes the binding of another. If the groups are on separate molecules, this could not occur, unless the molecules were to interact due to long-range forces—a type of interaction that has never been experimentally demonstrated. If positive interactions are observed, therefore, we know that the molecules involved must contain more than one combining group.

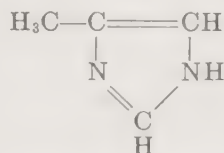
Binding by an Equivalent Set of Groups with Interactions between Them

We treat first the effect of interactions between the sites. The simplest case to consider is that which arises when the n binding sites on P are all equivalent to begin with, so that the probability that an A molecule will be bound is initially the same at all the sites. As soon as one such molecule is bound, however, the probability of binding at the neighboring sites is altered. Notable and relatively simple examples are given by the binding of many bases to certain types of metallic ions. The ion—for instance Cu^{++} , Fe^{++} , Zn^{++} , Co^{++} —here represents the molecule, P, and the ligand, A, is the base, which may be ammonia, pyridine, imidazole, or any one of a large variety of substances. The value of n for such systems is the characteristic coordination number for the central ion, being most commonly 4 for copper or zinc, and 6 for chromium, iron, cobalt, or nickel. The determination of binding in such systems and the evaluation of the association constants was first treated in a comprehensive fashion by Bjerrum (1941). As a concrete example of such a system we take the case of the binding of 4-methylimidazole to the ions of zinc and

copper (Nozaki *et al.*, 1957). Similar studies, with similar results, on unsubstituted imidazole were carried out earlier by Edsall *et al.* (1954a).



The 4-methylimidazolium ion (ImH^+)



4-Methylimidazole (Im)

The imidazole group is of great importance as a side chain of the histidine residue in proteins, and the methyl group of 4-methylimidazole is

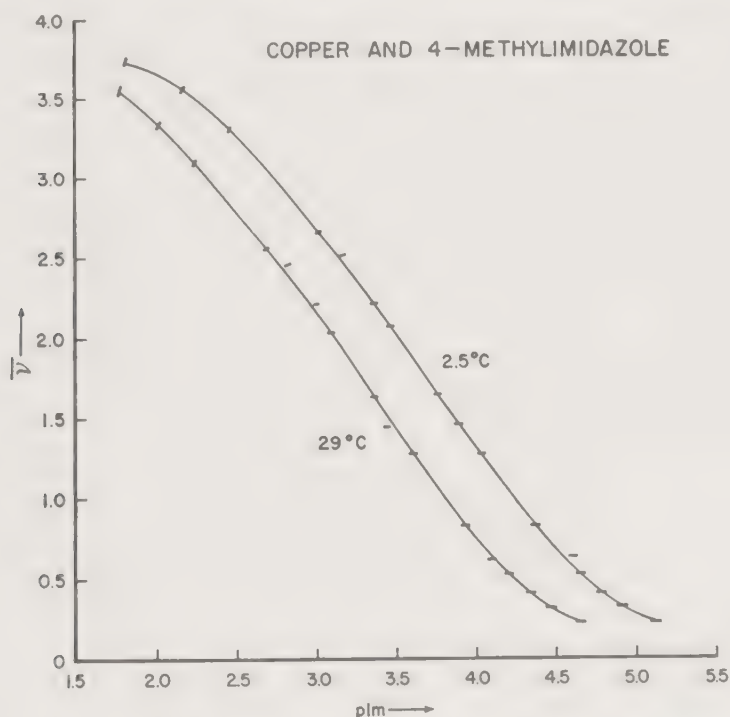


FIG. 8. The binding of 4-methylimidazole by Cu^{++} ions, as a function of the negative logarithm of the concentration of free basic imidazole (compare Fig. 4). (From studies by Y. Nozaki.)

located in the same position on the ring as the $-\text{CH}_2\text{CH}(\text{NH}_3^+)\text{COO}^-$ group which is attached to the imidazole ring in histidine. Therefore, although 4-methylimidazole is not itself observed in biochemistry, its study has important biochemical implications and it represents a system of relative simplicity. Its binding to the ions of copper and zinc has been studied by Y. Nozaki, and a plot of his results for these two ions in terms of the Bjerrum formation function is shown in Figs. 8 and 9.

The method of obtaining the data represented in Figs. 8 and 9 deserves some comment, since it is quite different from any of those described earlier. The basis of the procedure was first developed by Bjerrum in 1941. It requires that the pK' value of the imidazole be known under the conditions of temperature and ionic strength used in the actual binding studies with the other metallic ions. All experience shows that the base—in this case Im—is bound to the metal ion, but that the conjugate acid—in this case ImH^+ —is not. Hence if an ion such as Cu^{++} or Zn^{++} is added to a buffer mixture of Im and ImH^+ , the ratio $(\text{Im})/(\text{ImH}^+)$ is decreased by the addition, and the $p\text{H}$ is lower than it was in the buffer mixture before the metal ion was added. From the $p\text{H}$ shift the amount of Im bound may be computed (see equation 35 below),

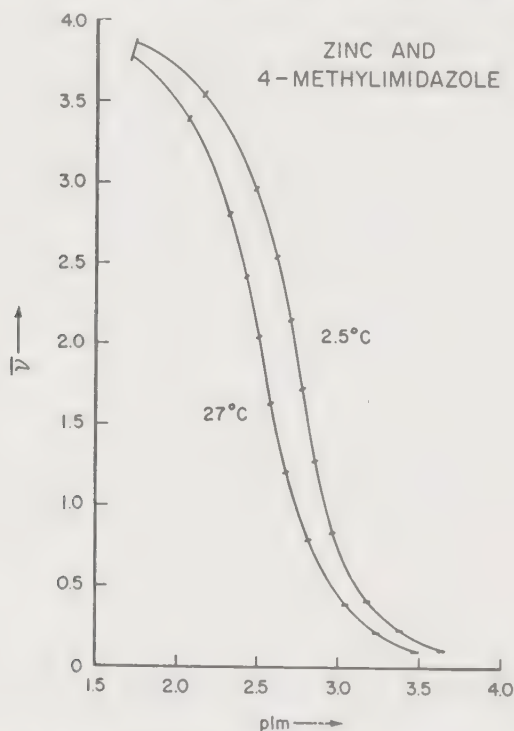


FIG. 9. The binding of 4-methylimidazole by Zn^{++} ions, plotted by the same relation used in Fig. 8. (From studies by Y. Nozaki.)

and \bar{v} may thus be evaluated as the ratio of the amount of Im bound to the total stoichiometric concentration of the metal ion added. Since the ligand $A = \text{Im}$ is uncharged, the complexes formed with a divalent ion, M^{++} , all carry the same charge and may be represented by the formulas MIm^{++} , MIm_2^{++} , and so forth.

The $p\text{H}$ decrease on addition of M^{++} is found to be much greater for copper than for zinc, showing clearly that the binding constants for the former must be considerably larger than for the latter. For quantitative determination of the binding, the concentration of metallic ion in the solution is maintained constant and the concentration of free imidazole available for reaction with the metallic ion adjusted by adding variable amounts of sodium hydroxide to solutions of 4-methylimidazolium nitrate. The copper and zinc ions are also added as the nitrates, since these are more nearly completely ionized than the chlorides.

In order to calculate the amount of imidazole bound to the copper or zinc ions,

it is necessary to know the total stoichiometric concentration, T_{Im} , of imidazole present in all forms, bound and unbound:

$$T_{\text{Im}} = (\text{Im}) + (\text{ImH}^+) + \sum_{i=1}^n i(\text{MIm}_i^{++}) \quad (31)$$

and to subtract from this the concentration of unbound basic imidazole (Im) and of imidazolium ion (ImH^+). The total stoichiometric concentration of metallic ion, in all forms, C_{M} , is

$$C_{\text{M}} = (\text{M}^{++}) + \sum_{i=1}^n (\text{MIm}_i) \quad (32)$$

and the total concentration of nitrate is

$$(\text{NO}_3^-) = C_{\text{M}} + T_{\text{Im}} \quad (33)$$

When Na^+OH^- is added to $\text{ImH}^+\text{NO}_3^-$, the cation ImH^+ is replaced by Na^+ ; (H^+) of course decreases, and (OH^-) increases. Given all these relations it is readily demonstrated from the requirements of electroneutrality that the concentration of ImH^+ is given by the equation

$$(\text{ImH}^+) = T_{\text{Im}} - (\text{Na}^+) + (\text{OH}^-) - (\text{H}^+) \quad (33.1)$$

The concentrations of both hydrogen and hydroxyl ions are generally small in such experiments compared with T_{Im} and (Na^+); the (OH^-) term generally is completely negligible, since the maximum value of hydroxyl ion concentration in the solutions studied by Nozaki is between 10^{-6} and 10^{-7} M , and (ImH^+) is experimentally chosen to be of the order of 10^{-1} . The (H^+) term is generally negligible for similar reasons. In most cases (ImH^+) can thus be taken as the total imidazole concentration, minus the amount of sodium hydroxide added to the original imidazolium nitrate solution; the amount of imidazole formed by dissociation of the imidazolium ion can be neglected except in the most acid solutions.

Since the value of pK' is known (see for instance Chapter 8, Table VI) and (ImH^+) is determined from equation (33.1), the concentration (Im) of free basic imidazole is directly determined from the pH measurements.

$$pH = pK' + \log \frac{(\text{Im})}{(\text{ImH}^+)} \quad (34)$$

The average number, $\bar{\nu}$, of imidazole molecules bound per metallic ion present is then given, from (24), (31) and (32), by the equation

$$\bar{\nu} = \frac{T_{\text{Im}} - [(\text{Im}) + (\text{ImH}^+)]}{C_{\text{M}}} \quad (35)$$

where C_{M} is defined by (32).

Inspection of Figs. 8 and 9 indicates that $\bar{\nu}$ approaches a limiting value of 4 at high concentrations of free Im, for both Cu^{++} and Zn^{++} . The probability that this is the true value of n is strengthened by the fact that studies with ammonia and numerous other ligands lead to the same value of 4 for the number of coordination positions available on

these ions for the binding of ligands. We therefore choose $n = 4$ as the basis of further calculations.⁴

Detectable binding is found at quite low concentrations of imidazole in the copper solutions, but only at appreciably higher concentrations in the zinc solutions. The curve rises much more steeply, however, in the zinc-imidazole than in the copper-imidazole system. The former is indeed considerably steeper than an ideal curve for a set of equivalent and independent groups (equations 12 and 13), and the latter is considerably flatter. This indicates immediately that the zinc curve involves positive interactions; the occupation of one coordination position on a zinc ion by Im increases the probability that another position on the same ion will also be occupied. The opposite is of course true in the copper system. We shall proceed to formulate these relations quantitatively in terms of the successive association constants.

The successive association constants involved in the reaction may be written as a special case of (24):

$$k_1 = \frac{(\text{MIm}^{++})}{(\text{M}^{++})(\text{Im})}; k_2 = \frac{(\text{MIm}_2^{++})}{(\text{MIm}^{++})(\text{Im})}; k_3 = \frac{(\text{MIm}_3^{++})}{(\text{MIm}_2^{++})(\text{Im})};$$

$$k_4 = \frac{(\text{MIm}_4^{++})}{(\text{MIm}_3^{++})(\text{Im})} \quad (36)$$

The intrinsic microscopic association constants, corrected for the statistical factor, are, from (26),

$$\kappa_1 = \frac{k_1}{4}; \kappa_2 = \frac{2k_2}{3}; \kappa_3 = \frac{3k_3}{2}; \kappa_4 = 4k_4 \quad (37)$$

The evaluation of the k 's or κ 's from data such as those illustrated in Figs. 8 and 9 may be carried out in various ways. An excellent discussion of several methods, with references to earlier work, is given by Hearon and Gilbert (1955).

Here we give only the fundamental principles of one such method,

⁴This does not mean that 4 is an absolute upper limit for the coordination of ligands by Cu^{++} and Zn^{++} . There is some evidence that a fifth, or even a sixth, coordination position may be occupied in the presence of certain ligands at high concentration. In the range of values of $p\text{Im}$ at which we are operating, however, we make a probably negligible error in ignoring them. The situation is similar in the acid titration of a protein, for instance. At $p\text{H}$ values about 1.5 to 2, the number of protons bound by the protein levels off to a fairly constant value, the "maximum proton binding capacity." There are, however, basic groups in proteins, such as the peptide linkages, or the $-\text{CONH}_2$ groups of asparagine and glutamine residues, which are still uncombined with protons at this $p\text{H}$. The basicity of these groups is so weak, however, that they do not bind protons appreciably until the $p\text{H}$ is well below zero. Hence perfectly consistent results are obtained by ignoring them at $p\text{H}$ values about 1.5.

which was developed by Scatchard, and first employed by Edsall *et al.* (1954a) in their work on imidazole. We define a function, Q , by the relation:

$$Q = \frac{\bar{\nu}}{(n - \bar{\nu})(A)} \quad (38)$$

where (A) , the ligand concentration, is in this case equal to (Im) .

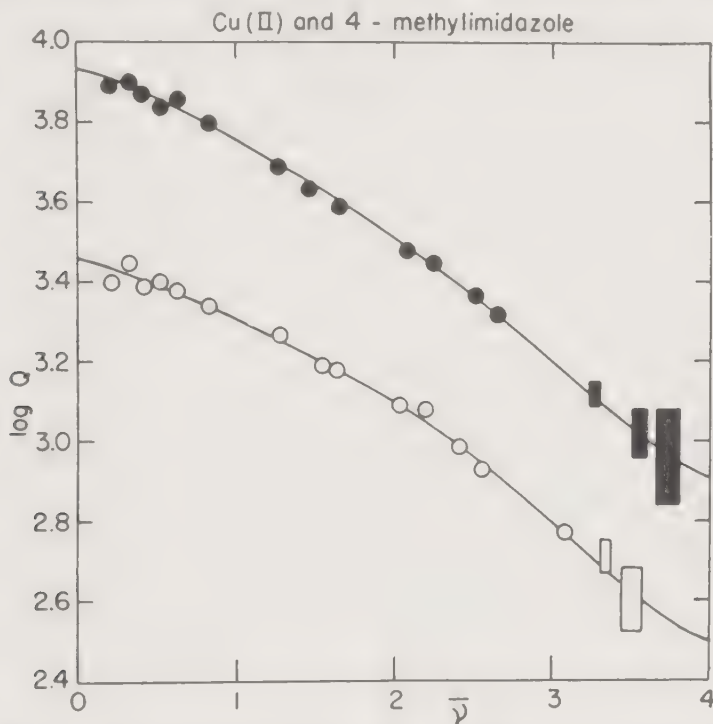


FIG. 10. The Scatchard function $\log Q$ plotted against $\bar{\nu}$ for the binding of 4-methylimidazole by Cu^{++} ion at two different temperatures. (From Nozaki, Gurd, Chen, and Edsall, 1957.)

On comparing (38) with (13), it is apparent that, for a set of equivalent and independent groups, Q is simply the intrinsic association constant, k , for this set of groups. A plot of Q (or $\log Q$) against $\bar{\nu}$ for such a system would be a horizontal straight line. The plot of the actual data, however, which is shown at two different temperatures for each of the two systems in Figs. 10 and 11, is neither horizontal nor linear. It is a curve with a negative slope for the copper-imidazole system. This is to be correlated with the fact that this system shows negative interactions between the binding sites on the copper ion. The zinc-imidazole system, which is characterized by positive interactions, gives a positive slope, as shown in Fig. 11.

In terms of the association constants k_1 to k_4 , inclusive, $\bar{\nu}$ is given by

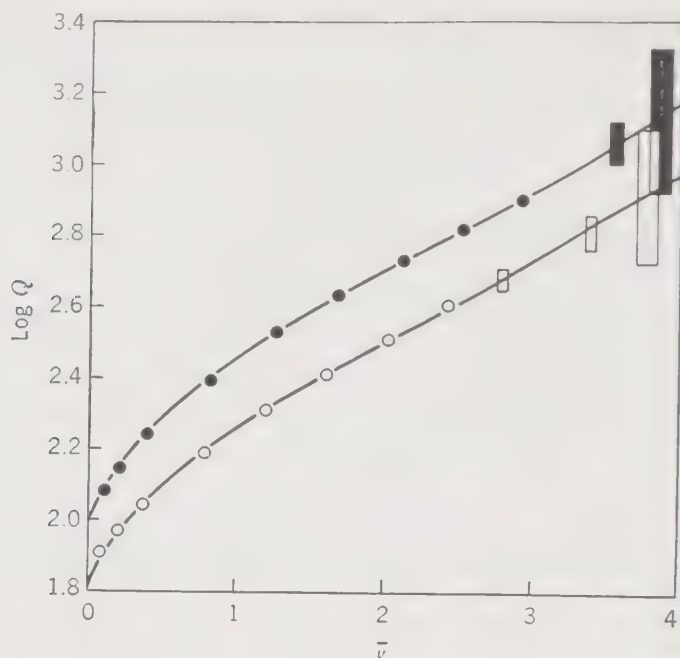


FIG. 11. $\log Q$ plotted against $\bar{\nu}$ for the binding of 4-methylimidazole by Zn^{++} ion at two different temperatures. (From Nozaki, Gurd, Chen, and Edsall, 1957.)

(24), for the special case in which $n = 4$. Or it may be given by the alternative expression in terms of the intrinsic constants κ_1 to κ_4 , inclusive:

$$\bar{\nu} = \frac{4[\kappa_1(A) + 3\kappa_1\kappa_2(A)^2 + 3\kappa_1\kappa_2\kappa_3(A)^3 + \kappa_1\kappa_2\kappa_3\kappa_4(A)^4]}{1 + 4\kappa_1(A) + 6\kappa_1\kappa_2(A)^2 + 4\kappa_1\kappa_2\kappa_3(A)^3 + \kappa_1\kappa_2\kappa_3\kappa_4(A)^4} \quad (39)$$

From (39) and the definition of Q we obtain

$$Q = \frac{\bar{\nu}}{(4 - \bar{\nu})(A)} = \frac{\kappa_1[1 + 3\kappa_2(A) + 3\kappa_2\kappa_3(A)^2 + \kappa_2\kappa_3\kappa_4(A)^3]}{1 + 3\kappa_1(A) + 3\kappa_1\kappa_2(A)^2 + \kappa_1\kappa_2\kappa_3(A)^3} \quad (40)$$

The limiting value of Q as (A) approaches zero is given from (40) by the relation

$$\lim_{(A) \rightarrow 0} Q = \frac{k_1}{4} = \kappa_1 \quad (40.1)$$

On the other hand, as (A) becomes infinite and $\bar{\nu}$ approaches 4, Q assumes the limiting form

$$\lim_{(A) \rightarrow \infty} Q = 4k_4 = \kappa_4 \quad (40.2)$$

Estimates of κ_2 and κ_3 may be derived from the limiting slope of the curve for $\log Q$ as a function of $\bar{\nu}$, as $\bar{\nu}$ approaches zero or 4, respectively, if κ_1 and κ_4 are already known. As $\bar{\nu}$ approaches zero the limiting slope of

the curve is given by

$$\lim_{\bar{v} \rightarrow 0} \frac{d \ln Q}{d \bar{v}} = \frac{3(\kappa_2 - \kappa_1)}{4\kappa_1} \quad (40.3)$$

and as \bar{v} approaches 4, and $1/(A)$ approaches zero, by

$$\lim_{\bar{v} \rightarrow 4} \frac{d \ln Q}{d \bar{v}} = \frac{3(\kappa_4 - \kappa_3)}{4\kappa_3} \quad (40.4)$$

On account of experimental uncertainties, the limiting slopes and the intercepts are both subject to significant experimental error. It is therefore generally necessary to estimate preliminary values of the k 's or κ 's from curves like those in Figs. 10 and 11, and from these values calculate curves such as those of Figs. 8, 9, 10, and 11. If the calculated curves differ significantly from the observed, the value of the k 's may be adjusted to give a better fit, and the process repeated. Some of this process of trial and error may be eliminated by the use of the mathematical procedures described by Hearon and Gilbert (1955). We shall not treat the further details of procedure here, however, since our main concern is with the results obtained.

The computed values of association constants for some systems involving imidazole, 4-methylimidazole, and ammonia are listed in Table I.

The contrast between the binding to zinc and that to copper is striking. The first molecule of 4-methyl-imidazole bound to copper is characterized by an intrinsic constant ($\log \kappa_1$) of 3.53 at 25°; the first one bound to zinc by a constant of only 1.84. The values of the successive constants steadily decrease for copper and increase for zinc, however, so that $\log \kappa_4$ for zinc (2.98) is actually greater than the corresponding value for copper (2.56). The values for unsubstituted imidazole are very similar to those for 4-methylimidazole. Thus the binding of one imidazole group to a copper ion discourages other imidazole molecules from combining, whereas the binding of one imidazole to a zinc ion offers a positive inducement to others to become linked also. This difference may be partially explained by the different spatial configuration of the copper and zinc complexes. The four Cu—N bonds in the former are probably arranged at the corners of a square; the four Zn—N bonds in the latter in the corners of a regular tetrahedron. If one attempts to construct a space model of these structures, the steric problems of placing four imidazole groups in the square configuration are considerably greater than in the tetrahedral. The difference between the two types of complexes appears to be more fundamental than this, however. The work of Bjerrum (1941) showed exactly the same sort of differences between the ammonia complexes of zinc and copper as are here shown between the imidazole complexes; and

TABLE I
SOME CONSECUTIVE ASSOCIATION CONSTANTS IN METAL-AMINE AND
METAL-IMIDAZOLE SYSTEMS
(Log κ values are given in parentheses)

System	Ionic strength	Temperature (°C)	Configuration	log k_1	log k_2	log k_3	log k_4	log k_5	log k_6
$\text{Ag}^{\text{I}} + 2\text{NH}_3$	2.0	30	Linear	3.20 (2.90)	3.83 (4.13)				
$\text{Hg}^{\text{II}} + 4\text{NH}_3$	*	22	Linear	8.8 (8.5)	8.7 (9.0)	1.00	0.78		
$\text{Zn}^{\text{II}} + 4\text{NH}_3$	2.0	30	Tetrahedral	2.37 (1.77)	2.44 (2.26)	2.50 (2.68)	2.15 (2.75)		
$\text{Zn}^{\text{II}} + 4$ imidazole	0.16	24	Tetrahedral	2.58 (1.98)	2.37 (2.19)	2.23 (2.41)	2.02 (2.62)		
$\text{Zn}^{\text{II}} + 4$ 4-methyl- imidazole	0.16	25	Tetrahedral	2.44 (1.84)	2.53 (2.35)	2.64 (2.82)	2.38 (2.98)		
$\text{Cd}^{\text{II}} + 6\text{NH}_3$	2.0	30	Tetrahedral	2.65 (2.05)	2.10 (1.92)	1.44 (1.62)	0.93 (1.53)	-0.4	-0.7
$\text{Cd}^{\text{II}} + 4$ imidazole	0.15	25	Tetrahedral	2.80 (2.20)	2.10 (1.92)	1.55 (1.73)	1.13 (1.73)		
$\text{Cu}^{\text{II}} + 5\text{NH}_3$	2	30	Planar	4.15 (3.55)	3.50 (3.32)	2.89 (3.07)	2.13 (2.73)	-0.5	
$\text{Cu}^{\text{II}} + 4$ imidazole	0.16	22.5	Planar*	4.36 (3.76)	3.57 (3.39)	2.85 (3.03)	2.06 (2.66)		
$\text{Cu}^{\text{II}} + 4$ 4-methyl- imidazole	0.16	25	Planar*	4.13 (3.53)	3.49 (3.31)	2.87 (3.05)	1.96 (2.56)		
$\text{Ni}^{\text{II}} + 6\text{NH}_3$	2.0	30	Octahedral	2.80 (2.02)	2.24 (1.84)	1.73 (1.60)	1.19 (1.32)	0.75 (1.15)	0.03 (0.81)
$\text{Ni}^{\text{II}} + 6$ imidazole	0.15	25	Octahedral	3.27 (2.49)	2.68 (2.28)	2.15 (2.02)	1.65 (1.78)	1.12 (1.52)	0.52 (1.30)

* Not specified.

Data for 4-methylimidazole from Nozaki *et al.* (1957). Other data taken from Table I of Gurd and Wilcox (1956). For data on imidazole see Edsall *et al.* (1954a). For ammonia complexes see Bjerrum (1941, 1950).

steric hindrance is certainly a minor factor for a ligand as small as ammonia. (See Table I.)

Indeed, as we have already indicated on p. 607, the existence of such positive interactions is characteristic of many important chemical and biochemical systems, as for instance in the successive binding of four oxygen molecules by hemoglobin, or in a very great number of oxidation and reduction processes, in which the positive interactions between successively bound or released electrons are often so great that two electrons appear to be added or removed simultaneously.

It is sometimes useful to analyze data such as those shown in Figs. 8 and 9 by another procedure which has been most explicitly described by Wyman (1948). This involves evaluating $\log [\bar{\nu}/(n - \bar{\nu})]$ as a function of $\log (A)$. For a set of equivalent and independent groups, the resulting curve is a straight line of unit slope.

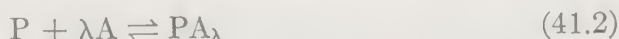
Even where there are strong interactions, however, such a curve is often found to be nearly straight over a considerable portion of the middle range. Whenever this is so, equation (24) may be replaced by a simpler approximate equation which often proves useful. The slope of the curve for $\log [\bar{\nu}/(n - \bar{\nu})]$ as a function of $\log (A)$ may be denoted by

$$\frac{d}{d \log (A)} \log \frac{\bar{\nu}}{n - \bar{\nu}} = \frac{\partial \ln [\bar{\nu}/(n - \bar{\nu})]}{\partial \ln (A)} = \lambda \quad (41)$$

Then, to the extent to which λ is independent of (A) , it follows that

$$\frac{\bar{\nu}}{n - \bar{\nu}} = k'(A)^\lambda \quad (41.1)$$

Here k' , the constant of integration, may be interpreted, formally at least, as the equilibrium constant for the apparent reaction



For $\lambda = 1$, equation (41.1) of course reduces to (12.1). In general λ is given from (24) and (41) by the equation

$$\begin{aligned} \lambda &= \frac{\partial \ln [\bar{\nu}/(n - \bar{\nu})]}{\partial \ln (A)} = \frac{n}{\bar{\nu}(n - \bar{\nu})} \frac{d\bar{\nu}}{d \ln (A)} \\ &= \frac{n}{\bar{\nu}(n - \bar{\nu})} \frac{d \left[\sum i L_i^*(A)^i / \sum L_i^*(A)^i \right]}{d \ln (A)} \\ &= \frac{n}{\bar{\nu}(n - \bar{\nu})} \left[\frac{\sum i^2 L_i^*(A)^i}{\sum L_i^*(A)^i} - \left(\frac{\sum i L_i^*(A)^i}{\sum L_i^*(A)^i} \right)^2 \right] \\ &= \frac{n}{\bar{\nu}(n - \bar{\nu})} [\bar{\nu}^2 - (\bar{\nu})^2] \end{aligned} \quad (41.3)$$

This equation is quite general for any system in which the L^* values may be regarded as true constants and is independent of the approximation involved in deriving equation (41.1).⁵

Thus the slope of this curve is proportional to the standard deviation of ν at any point—that is, to the difference between the mean square value of ν and the square of the mean value. The derivation is practically identical with that of the Linderstrøm-Lang equation (Chapter 9, equation 49) for the slope of a titration curve.

It is often of particular interest to know the slope of such a curve at the mid-point—that is, when $\bar{\nu} = n/2$. From (41.3) this is

$$\lambda_{\text{mid}} = \frac{1}{n} [\bar{\nu^2} - (\bar{\nu})^2] = \frac{1}{n} \left(\bar{\nu^2} - \frac{n^2}{4} \right) = \frac{4\bar{\nu^2}}{n} - n \quad (41.4)$$

Since $(\bar{\nu^2} - \bar{\nu}^2) \geq 0$ for any distribution of ν values, corresponding to a given $\bar{\nu}$, among the various molecules of P in the system, λ is always positive. For a random distribution among the binding sites—that is, in the absence of interactions— $\lambda = 1$, as we have already mentioned. If $\lambda_{\text{mid}} > 1$, this is a definite indication of positive interactions among the binding groups; if $\lambda_{\text{mid}} < 1$, negative interactions are generally implied although some positive interactions may exist in certain systems, even when $\lambda_{\text{mid}} < 1$.

The fact that—if interactions are positive—the standard deviation of $\bar{\nu}$ is greater than that for a random distribution, and hence that $\lambda_{\text{mid}} > 1$, is qualitatively easy to appreciate. Consider for instance, in the zinc-imidazole system, the distribution of imidazole molecules among the zinc ions when $\bar{\nu} = 2$. The fact that positive interactions exist means that, if one site on a zinc ion chosen at random is found to be occupied by an imidazole molecule, then there is a better than even chance of finding the neighboring sites on the same ion also occupied. Thus there will be more ions in the solution with $\bar{\nu} = 3$ or 4, and also more with $\nu = 0$ or 1, than would be expected for a random distribution corresponding to $\bar{\nu} = 2$. This of course is just another way of saying that the standard deviation of ν is higher than would be expected for a random distribution.

In Fig. 12 the data of Figs. 8 and 9 are plotted with $\log [\bar{\nu}/(4 - \bar{\nu})]$ as ordinate and $\log (Im)$ as abscissa. The great difference between the curves for copper and zinc is again apparent: λ_{mid} for Cu^{++} is 0.6, for Zn^{++} 2.0. The difference between the values of $\log (Im)$ at the mid-point, for the Cu^{++} and Zn^{++} curves, gives a measure of the value of $-\log k'$, where k'

⁵ Equation (41.3) is identical with equation (4) on p. 439 of the review by Wyman (1948), although somewhat different in appearance.

is the "apparent affinity constant" defined in (41.1) and (41.2). This constant is obviously larger for Cu^{++} than for Zn^{++} .

The type of plot used in Fig. 12 has been much used in the study of the equilibrium between oxygen and hemoglobin, for which $n = 4$ as in the cases just discussed. The abscissa in this is $\log (\text{O}_2)$, where (O_2) is the molar concentration of free O_2 in the liquid, or $\log p$, where p is the partial pressure of oxygen in the system. Hemoglobin, in binding oxygen,

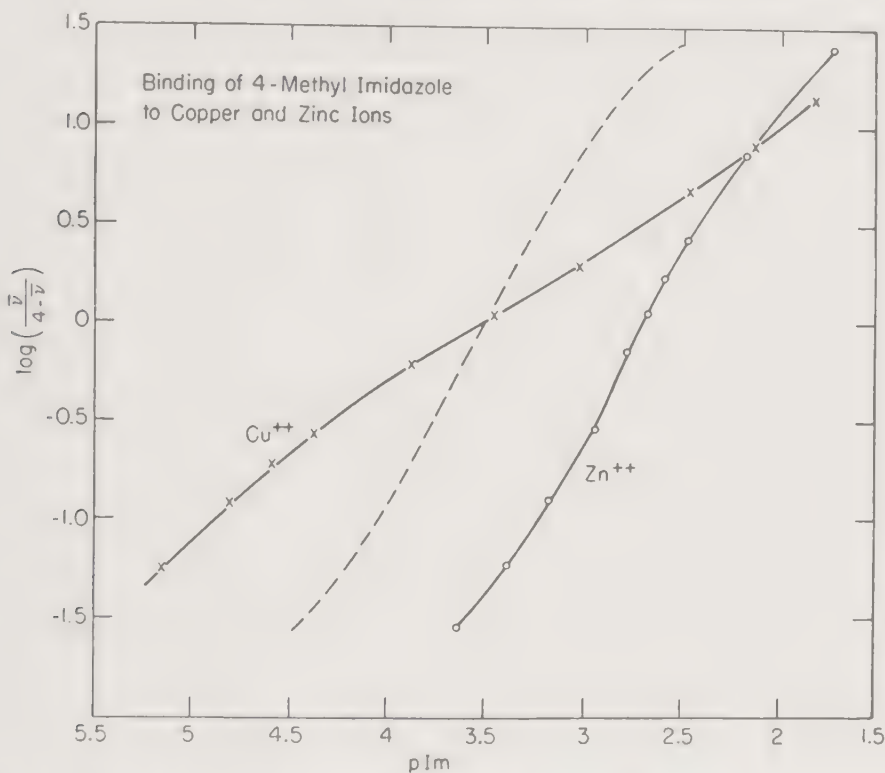


FIG. 12. Plot of $\bar{v}/(4 - \bar{v})$ as a function of the logarithm of the concentration of free basic 4-methylimidazole, using the same data shown in Figs. 8, 9, 10, and 11. The dotted line shows the Zn^{++} curve displaced parallel to itself so that its mid-point coincides with that of the Cu^{++} curve.

shows very strong positive interactions, the value of λ_{mid} being 2.8. Interactions in the hemoglobin system are discussed in detail in Volume II.

The analysis of binding curves in terms of λ_{mid} and the value of (A) at the mid-point does not, of course, lead directly to the evaluation of all the individual association constants. It serves, however, as a simple way of characterizing the system in terms of a mean affinity constant and an interaction coefficient. It is most useful in systems with strong interactions in which the successive k values are quite close together. If the k 's are spread far apart, as for instance in the proton binding of such dibasic

acids as glycine, then λ_{mid} is nearly zero and cannot be used to give reliable information.

Complexes of Biochemical Substances with Metallic Ions: Some General Considerations

The binding of 4-methylimidazole to copper or zinc ions has been considered in some detail as a relatively simple example of an interaction between certain metallic ions and an organic ligand. We now consider some of the more general relations involved. The particular example just discussed involved ions with a coordination number of 4. For each type of metallic ion there is generally such a number, and in the great majority of cases it is either 2, 4, or 6. In the special case of H^+ ion, it is 1. The usual value⁶ is 2 for Hg^{++} , and for Cu^+ , Ag^+ , and Au^+ ; such complexes are generally linear, e.g., $\text{Cl}-\text{Hg}-\text{Cl}$. The value of 4 is characteristic for Cu^{++} , Zn^{++} , Cd^{++} , Be^{++} and some other ions; values of both 4 and 6 are found for Mg^{++} , Mn^{++} , Fe^{++} , Co^{++} , and Ni^{++} ; and 6 is almost invariably characteristic of Cr^{+++} , Fe^{+++} , and Co^{+++} . Symmetrical 4-coordinate complexes are generally square, as is common with Cu^{++} complexes, or tetrahedral, as is common with Zn^{++} complexes. In 6-coordinate complexes, the ligands are located at the corners of an octahedron. The extension of equations (32), (33), and (35) to (41), inclusive, to coordination numbers higher than 4 is fairly obvious, and need not be discussed farther here. We note, however, that for an ion with a characteristic coordination number, n , it is sometimes useful to define a "mean association constant," $\bar{K} = (k_1 k_2 \cdots k_n)^{1/n}$ to express briefly its affinity for a given ligand. The use of this constant is illustrated in the discussion below.

Some values of association constants with metallic ions, for compounds of biochemical or general interest, are listed in Table II. Only a few among the great multiplicity of data available are listed in this table. Many others may be found in the reviews by Bjerrum (1950), Martell and Calvin (1952), Williams (1953), Schwarzenbach (1954), and Gurd and Wilcox (1956). From all these data some important general relations emerge.

1. The affinities of many basic ligands for certain metallic ions run roughly parallel to their affinities for protons. Bjerrum has formulated

⁶ The mercuric ion, Hg^{++} , can form with halogen ions or ammonia molecules a series of four complexes, e.g. with chloride ion HgCl^+ , HgCl_2 , HgCl_3^- , and HgCl_4^{--} . The values of $\log k_1$ and $\log k_2$ for the formation of these complexes (see 24) are near 7; those of $\log k_3$ and $\log k_4$ near 1. The two latter are so small compared to the first two that they may be neglected for many purposes, and we consider the characteristic coordination number of mercury as 2. (For further details on mercury-halogen association constants, see Sillén, 1949.) Similar remarks apply to some metals with a coordination number of 4; see footnote 4 above.

TABLE II
ASSOCIATION CONSTANTS FOR THE COMBINATION OF VARIOUS METAL IONS
WITH ORGANIC BASES

A. Combination with Glycinate Ions

Metal ion	$\log k_1$	$\log k_2$	$\log k_3$	Ionic strength	Temperature (°C)
Ba ^{II}	0.77	*	*	0.0	25
Ca ^{II}	1.43	*	*	0.0	25
Mg ^{II}	3.44	*	*	0.0	25
Mn ^{II}	3.44	*	*	0.0	25
Co ^{II}	4.61	3.75	2.56	0.5	20
Ni ^{II}	5.77	4.80	3.61	0.5	20
Cd ^{II}	3.88	3.18	1.92	0.5	20
Zn ^{II}	4.80	4.14	2.54	0.5	20
Cu ^{II}	8.22	6.97	*	0.5	20
Pb ^{II}	5.47	4.02	*	0.0	25
Hg ^{II}	10.3	8.9	*	0.5	20
Ag ^I	3.7	3.3	*	0.5	20

B. Comparison of Association Constants ($\log k_1$) for Combination with Citrate, Oxalate, Glycinate, and Ethylenediamine

Metal ion	Citrate ⁻⁻⁻	Oxalate ⁻⁻	Glycinate ⁻	Ethylenediamine
Ba ^{II}	3.0	2.33	0.77	†
Ca ^{II}	3.2	3.00	1.43	†
Mg ^{II}	3.2	3.43	3.44	†
Mn ^{II}		3.89	3.44	2.73
Co ^{II}		4.70	5.23; 4.61	5.89
Ni ^{II}		5.3	6.18; 5.77	7.52
Zn ^{II}		4.89	5.52; 4.80	5.71
Cu ^{II}		6.16	8.62; 8.22	10.55

* The data were interpreted without the necessity of assuming that such a step occurs to a measurable extent.

† The affinity of these metal ions for amines is very small.

Data for citrate from Martell and Calvin (1952). Other values from Gurd and Wilcox (1956, Tables III and IV).

Where two values of $\log k_1$ are given for glycinate⁻, the one on the left is measured under comparable conditions to that determined for oxalate⁻⁻; the one on the right is measured under comparable conditions to ethylenediamine.

this relation by comparing the mean association constants for metallic ions, defined above and expressed here as $\log \bar{K}$ values, with the proton affinities expressed as $\log K_{H^+} = pK_A$ for the conjugate acids. Thus Bjerrum found for a large number of nitrogenous bases, including ammonia, methylamine, aniline, benzylamine, ethanolamine, glycinate (amino-

acetate) ion, and many others, that the ratio of $\log \bar{K}$ for Ag^+ ion to pK_A was given by $\log \bar{K}_{\text{Ag}^+}/pK_A = 0.37 \pm 0.03$, over a range of pK_A values from 2 to 11. The corresponding ratio for several secondary aliphatic amines was near 0.29, and for some tertiary amines near 0.23. (The ratio for pyridine, however, was like that of a primary amine.) The corresponding ratio for mercuric complexes, $\log \bar{K}_{\text{Hg}^{++}}/pK_A$, was between 0.7 and 1

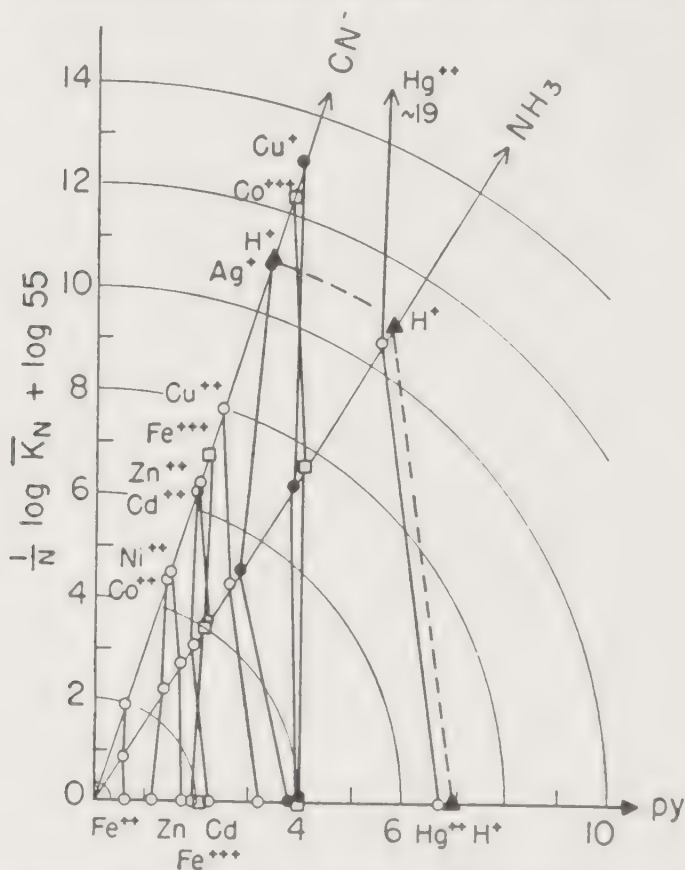


FIG. 13. The binding of cyanide ion, ammonia, and pyridine by hydrogen ion and by various metallic ions. Since the molar concentration of water is approximately 55, the term $\log 55$ is added to the ordinate, so that the constants given are for the displacement of water by the three ligands, when attached to the indicated cations. Angles, measured from the horizontal, of 58° for the NH_3 line and 72° for the CN^- line were chosen to give the best agreement in projections on the line for py (pyridine). (From C. D. Coryell, in Gurd (1954), p. 97.)

for a number of bases, and the ratios for Cu^{++} complexes, $\log \bar{K}_{\text{Cu}^{++}}/pK_A$, also were found to lie in the same range for many bases, including carboxylate ions as well as nitrogenous bases. For mercuric ion, however, this relation breaks down completely for the conjugate base (RS^-) of a sulfhydryl compound. The pK_A value of a simple mercaptan is near 10, but the value of $\log \bar{K}_{\text{Hg}^{++}}$ for RS^- is over 20.

2. The affinity of a given metallic ion for different bases often falls in a series which is the same for several different metals. Thus Bjerrum (1950) stated such a relationship in terms of the quantity $\log \bar{K} + \log 55$. The number 55 denotes the molality of water in the system, and the addition of the term $\log 55$ allows for the fact that the added basic ligand is competing with water molecules for a coordinate bond on the metallic ion. Bjerrum stated the quantitative rule that $(\log \bar{K} + \log 55)$ was in the ratio 1.7:1:0.6 for the ligands CN^- , NH_3 , and pyridine in their association with various metals. This relation is illustrated graphically in Fig. 13. The experimental values for the H^+ ion, especially in HCN , deviate from Bjerrum's relation, but for the most part there is good agreement. Similar correlations, generally much less exact, are often found in other cases, and they frequently furnish a useful guide for estimating association constants in cases in which they have not been experimentally measured.

TABLE III
LOGARITHMS OF ASSOCIATION CONSTANTS FOR AMMONIA AND ETHYLENEDIAMINE
WITH CERTAIN METALLIC IONS
(Temperature 25°, ionic strength 1.0)

Substance		Co^{++}	Ni^{++}	Ca^{++}	Zn^{++}	Cd^{++}
Ammonia						
(A)	$\log [(\text{MA}_2^{++})/(\text{M}^{++})(\text{A})^2] = \log k_A$	3.68	4.99	7.68*	4.69	4.69
Ethylene-						
diamine	$\log [(\text{MEn}^{++})/(\text{M}^{++})(\text{En})] = \log k_{\text{En}}$	5.99	7.77	10.73*	5.92	5.63

* Ionic strength 1.3.

From G. Schwarzenbach (1954).

3. Compounds containing two or more functional groups, which can act as ligands and thereby form chelate rings, show much stronger binding than monovalent compounds containing the same functional group. This may be illustrated by some comparative data for the binding of ammonia and of ethylene-diamine by certain divalent metallic cations (M^{++}), listed in Table III. The total energy change, and therefore the heat of reaction, ΔH , involved in the binding of two NH_3 molecules to such a cation should not be very different from the binding of one molecule of $\text{H}_2\text{N}\cdot\text{CH}_2\cdot\text{CH}_2\cdot\text{NH}_2$, since the bonds formed are of very similar character.⁷ It is plain from Table III, however, that the *free energy*

⁷ We remind the reader that the reaction is not a simple association, but rather a displacement of H_2O molecules, in the coordination shell around the cation, by NH_3 molecules or NH_2 groups. This, however, does not affect the validity of the point made in the text.

changes for the two processes are quite different. In view of the fundamental thermodynamic relation (see Chapter 4):

$$\Delta F^\circ = -RT \ln k = \Delta H^\circ - T\Delta S^\circ$$

it is plain that, since $\ln k_A$ is much less than $k_{E\cdot}$, and since ΔH° should not be very different for the two, the greater association constants for the ethylenediamine complexes must be due to the entropy term above. That is, the value of $\Delta S_{E\cdot}^\circ$ is more positive than ΔS_A° . In terms of the relations between entropy and probability (p. 224), this difference is readily understandable, at least qualitatively. If one NH_2 group of ethylenediamine has become linked to the ion M^{++} , then the probability is great that the second group is close enough to react also, since the two groups are linked together, and the distances and bond angles are such that a five-membered ring is readily formed, involving M^{++} and ethylenediamine. On the other hand, the fact that one NH_3 molecule has become bound to M^{++} gives no guarantee whatever that a second molecule will be available to react, since these molecules are quite independent. This line of thought has been further developed by Schwarzenbach (1954), whose discussion is highly recommended. Schwarzenbach points out, with illustrative data, that five-membered chelate rings are generally the most stable; six-, seven-, and eight-membered rings are progressively less so.

4. For nitrogenous ligands, such as ammonia, amines, and imidazoles, the first association constants with divalent metallic ions of the first transition series increase in the order $\text{Mg}^{II} < \text{Mn}^{II} < \text{Fe}^{II} < \text{Co}^{II} < \text{Ni}^{II} < \text{Cu}^{II}$. For Zn^{II} , however, there is an abrupt fall in affinity, well below the value for Cu^{II} . Some of these relations are illustrated in Tables I and II.

Carboxylate ions generally show distinctly lower affinity for all these metallic ions than is shown by nitrogenous groups, although the order of affinity constants is commonly the same as that listed above. Carboxyl ions, however, especially those of hydroxy acids, such as citrate ion, bind the ions of the alkaline earths more strongly than do the nitrogenous groups. The order of increasing affinity constants is generally $\text{Ba}^{II} < \text{Ca}^{II} < \text{Mg}^{II}$. In this respect, as in others, Pb^{II} behaves like the alkaline earths.

These differences are important in the interactions of various ions with proteins. The work of Gurd and Goodman (1952), for instance, carried out by the equilibrium dialysis technique, shows that serum albumin between pH 5 and 6.8 binds Zn^{II} ions almost entirely through the histidine imidazole groups, with an intrinsic binding constant, $\log k = 2.76$, for each of the 16 or 17 imidazole groups of the albumin molecule. This is comparable to the values of $\log \kappa$ for Zn^{II} with imidazole

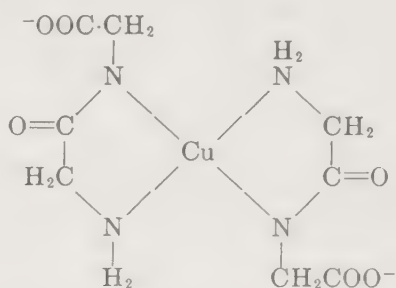
or 4-methylimidazole (Table I). There is of course competition between H^+ ions and Zn^{II} ions for the imidazole nitrogens. By contrast Pb^{II} (Gurd and Murray, 1954) binds primarily to the carboxylate groups of human serum albumin. In 0.4 M lead perchlorate and at pH 4.6, 73 Pb^{II} ions were bound per mole of albumin, and this was certainly not a maximum figure. Competition between lead and hydrogen ions was observed in the pH range of carboxyl dissociation, between 3 and 5.7. The Pb^{II} ions, when many were bound, caused precipitation of the protein, but Gurd and Murray produced good evidence for believing that the conditions were comparable to those in solution. The protein redissolved, apparently quite undenatured, after removal of the Pb^{II} ions with a chelating agent.

Gurd and Murray further studied the simultaneous binding of Pb^{II} and Zn^{II} in serum albumin solutions containing both these ions. At pH 5.7 and 0° , the amount of Pb^{II} bound was the same as in the solutions in which no Zn^{II} was present, and the binding of Zn^{II} also appeared to be unaffected by the presence of Pb^{II} . Therefore it appeared clear that the two ions were combining at different sites on the protein molecule. In contrast, the addition of cadmium (Cd^{II}) ions, with Zn^{II} also present, showed that the bound Zn^{II} diminished as the bound Cd^{II} increased. In this case both ions were evidently competing for imidazole groups, as might be expected because of their similar electronic configuration, both belonging to group IIB in the periodic table. For a thoughtful discussion of the whole situation, see Gurd and Wilcox (1956).

5. Simple peptides combine less firmly with most ions to form chelate complexes than do most α -amino acids. Thus the glycylglycinate anion combines with Cu^{II} , with $\log k_1 = 6.04$, as compared with 8.2 for the glycinate anion (see Table II). There is evidence (Dobbie and Kermack, 1955; Datta and Rabin, 1956) that Cu^{II} links to the glycylglycinate anion through the terminal amino group, and the nitrogen of the peptide linkage, the carboxylate ion not being involved.⁸ This of course involves an ionization of the hydrogen of the $-CO \cdot NH-$ group; the structure of the

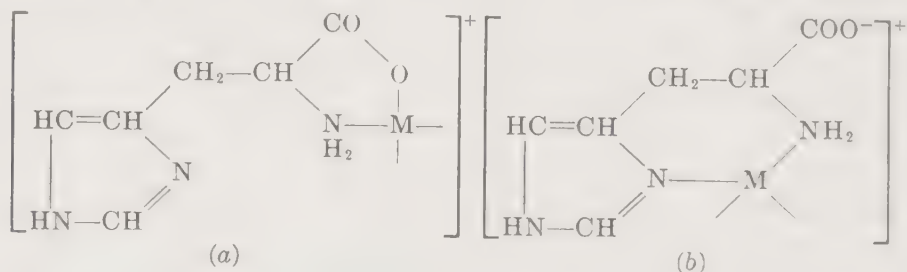
⁸ Actually the situation may be more complex than this (Rabin, 1956). The cobaltous and manganous complexes of dipeptides apparently involve the amino nitrogen and the peptide oxygen. The amino nitrogen-metal ion interaction involves a coordinate bond which is primarily covalent, whereas the oxygen-metal ion interaction is primarily electrostatic (Sen *et al.*, 1955). Rabin (1956) concludes that the cupric complexes may form before the ionization of the peptide hydrogen has occurred, and that in this case their structures are identical to those of the cobaltous and manganous complexes. When the peptide hydrogen ionizes, a rearrangement may occur giving a complex which involves coordination of the peptide nitrogen. It is even possible to have a threefold coordination involving the ionized carboxyl group of the peptide as well as the amino group and the peptide linkage. The situation is not completely clarified, and more work is needed.

complex $[(\text{Gly}\cdot\text{Gly})_2\text{Cu}]^{--}$ is probably



Here we have preferred to assume the *trans* configuration of the two glycylglycine anions, as suggested by Gurd and Wilcox (1956), rather than the *cis* configuration proposed by Dobbie and Kermack (1955). The *trans* arrangement tends to minimize electrostatic repulsions by placing the ionized ---COO^- groups as far apart as possible.

Amino acids with a third reactive group, such as histidine, may form complexes with metallic ions in several possible ways. In the case of histidine, the favored form of chelate complex, with ions such as Cu^{II} , is probably one involving the α -amino group and a nitrogen of the imidazole ring, to form a six-membered chelate ring (b), rather than the usual five-membered ring involving the ---COO^- and the α -amino groups (a):



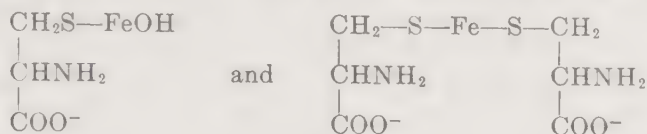
Still other structures are possible. A more extensive discussion of other polyfunctional amino acids has been given by Gurd and Wilcox (1956).

6. Thiol compounds occupy a special place in their reactions with metallic ions. The association constant, k_{RSM} ,

$$k_{\text{RSM}} = \frac{(\text{RSM})}{(\text{RS}^-)(\text{M})}$$

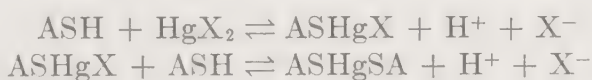
is greater than 10^{20} if M is Hg^{II} , about 10^{19} for Cu^{I} , 10^{15} for Ag^{I} , 10^{11} for Pb^{II} , 10^7 for Zn^{II} and Cd^{II} , and considerably less for nearly all other ions (see Gurd and Wilcox, 1956, p. 351, for a table of data; detailed studies of the interaction of cysteine, glutathione, and other thiol compounds with Hg^{II} and Cu^{I} have been made by Stricks and Kolthoff, 1953, and by Stricks *et al.*, 1954). Cysteine in alkaline solution forms soluble com-

plexes with Fe^{II} and Fe^{III} ions; according to Tanaka *et al.* (1955), the following complexes with Fe^{II} (and perhaps others) exist at pH 10 to 11:



and Fe^{III} complexes with the formulas $\text{FeOH}(\text{RS})_2^{--}$ and $\text{Fe}(\text{RS})_3^{---}$ were demonstrated. Iron complexes with sulfhydryl compounds have aroused much biochemical interest in the past, since the so-called "autoxidation" of cysteine in mildly alkaline solution depends on the formation of complexes with traces of heavy metal ions (usually iron) present in ordinary water even after distillation (see, for instance, Michaelis, 1946).

Similar reactions presumably often occur with sulfhydryl groups in proteins, including enzymes. This field of study is so vast that we merely mention it in passing, referring the reader to the review by Putnam (1953), which gives an extensive discussion, with references. One system of special interest may be mentioned here, however—namely, serum mercaptalbumin, which contains a single sulfhydryl group in a molecule of molecular weight near 66,000. Human mercaptalbumin was first separated and crystallized from total human serum albumin by Hughes (1947) as a mercury derivative containing two molecules of albumin per atom of mercury. He showed (see also Hughes, 1950) that a mercury dimer of mercaptalbumin, with two albumin molecules linked by Hg^{II} through their sulfhydryl groups, also exists in solution, in a reversible equilibrium with mercaptalbumin (ASH) and its simple mercury derivative, according to the reactions



Here X is generally chloride or acetate. The mercury dimer, ASHgSA , may be dissociated by adding another equivalent of HgX_2 into two moles of ASHgX . All these reactions are reversible. The velocity and equilibrium constants of these reactions have been extensively studied, largely by means of light-scattering measurements (Edelhoch *et al.*, 1953; Edsall *et al.*, 1954b; Edsall, 1954; Kay and Edsall, 1956).

Electrostatic Effects on Binding of Ions

If ions are bound by a macromolecule, P, each ion bound alters the net charge on P in such a way that the electrostatic work of adding another ion of the same kind is increased. If P can be treated as a charged

sphere with the mean net charge, \bar{Z} —expressed in proton units—spread uniformly over the surface of the sphere, then the treatment is identical with that already outlined in Chapter 9, pp. 512–522, inclusive. The most detailed studies of such electrostatic interactions are those involving the binding of protons, already discussed in Chapter 9. An additional important class of such interactions is presented by the binding of anions to

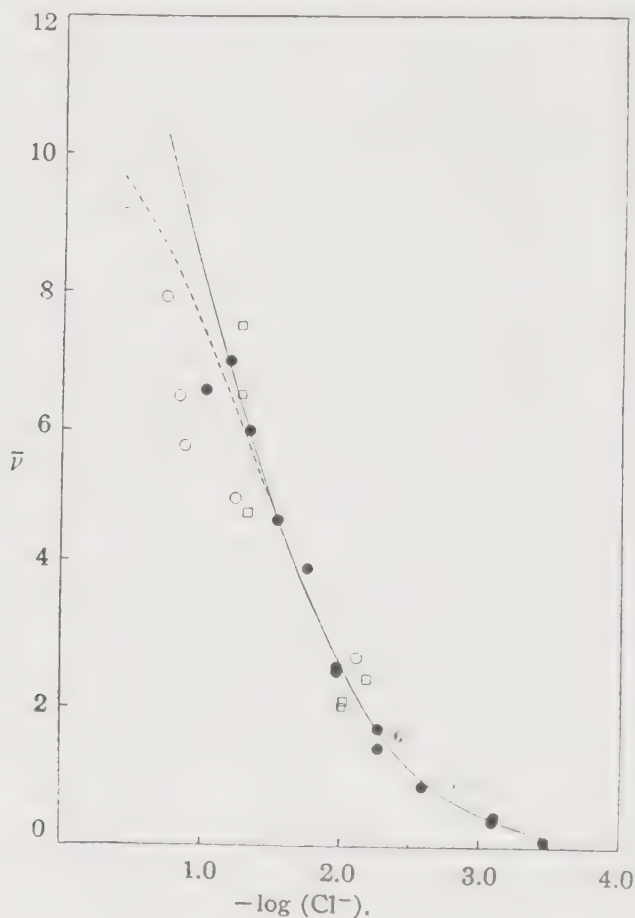


FIG. 14. Binding of chloride ions by human serum albumin, plotted by the method of Fig. 4. (From Scatchard, Scheinberg and Armstrong, 1950.)

serum albumin. The anions which have been most thoroughly studied in this respect are chloride and thiocyanate (Scatchard *et al.*, 1950). The methods employed to determine the binding of these ions were equilibrium dialysis and electromotive force determination on isoionic human serum albumin, with electrodes reversible to Cl^- and SCN^- ions; these methods have already been discussed (pp. 594–598). The results, expressed as plots of the formation functions, are shown in Figs. 14 and 15. It is immediately apparent from these data that thiocyanate is more

strongly bound than chloride. Thus $\bar{\nu}_{\text{Cl}^-}$ is 4 when $-\log (\text{Cl}^-)$ is near 1.5; but $\bar{\nu}_{\text{SCN}^-}$ is 4 when $-\log (\text{SCN}^-)$ is near 2.8. Likewise the binding of chloride appears to be leveling off toward a value of $\bar{\nu}_{\text{Cl}^-}$ near 10, whereas $\bar{\nu}_{\text{SCN}^-}$ obviously attains a value of 35 or more at high values of (SCN^-) . In the studies of Scatchard *et al.*, sodium chloride and thiocyanate were

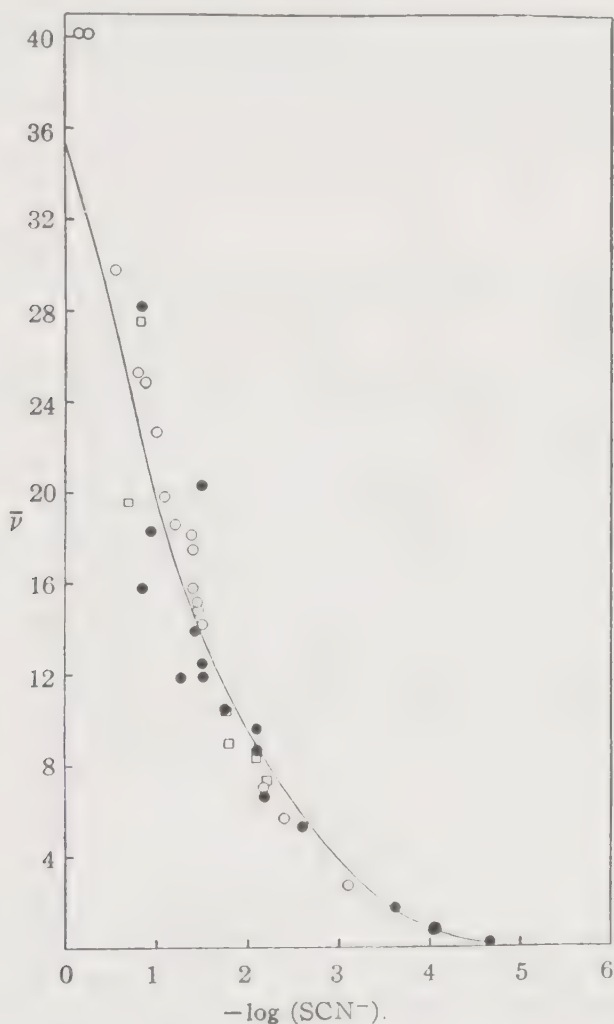


FIG. 15. Binding of thiocyanate ions by human serum albumin. (From Scatchard, Scheinberg, and Armstrong, 1950.)

added to the isoionic protein, so that \bar{Z} is zero, to a close approximation, in the absence of added salt, and $\bar{Z} = -\bar{\nu}$ in the presence of salt, since the bound ions are negative and univalent.

To calculate the electrostatic effects on binding, assuming the albumin molecule to be like a uniformly charged sphere, we may try the use of equation (63) of Chapter 9, remembering that Cl^- and SCN^- are both

univalent anions. Hence $Z_i = -1$ for both. If we write simply \bar{Z} for \bar{Z}_p , the equation then becomes

$$\bar{\nu} = \frac{nk(A)e^{2w\bar{Z}}}{1 + k(A)e^{2w\bar{Z}}} \quad \text{or} \quad \frac{\bar{\nu}}{n - \bar{\nu}} = k(A)e^{2w\bar{Z}} \quad (42)$$

Here n is again the number of reactive sites on the albumin, k is the intrinsic binding constant, w is the parameter defined in Chapter 9, equation (53), and (A) denotes the activity of (Cl^-) or (SCN^-) . Equation (42) can be rearranged to give

$$\frac{\bar{\nu}e^{-2w\bar{Z}}}{(A)} = k(n - \bar{\nu}) \quad (43)$$

Since we are dealing with isoionic protein, we can write $e^{-2w\bar{Z}} = e^{2w\bar{\nu}}$. Except for the exponential term, (43) is identical with (16), and the data of Figs. 14 and 15 can be examined in terms of a plot similar to that of Fig. 6. As in that figure the abscissa is $\bar{\nu}$, but the ordinate is $\bar{\nu}e^{-2w\bar{Z}}/(A)$. If the assumptions made in deriving (42) and (43) hold, the data should fall on a straight line; and as in Fig. 6 the intercept on the abscissa would then be $\bar{\nu} = n$, and that on the ordinate would be $\bar{\nu}e^{-2w\bar{Z}}/(A) = kn$.

The data of Fig. 15, plotted in this fashion, are shown in Fig. 16. It is obvious at once that the curve is not linear, and this indicates that more than one kind of binding group on the albumin molecule is involved. We may try to deal with this situation by the assumption that there are two or more classes of groups, the n_i groups in the first class having the intrinsic constant k_i , and so forth. In this case we can extend equation (42) by writing $\bar{\nu}$ as the sum of a series of terms similar to those in equation (12.1):

$$\bar{\nu} = \sum_{i=1}^m \frac{n_i k_i (A) e^{2w\bar{Z}}}{1 + k_i (A) e^{2w\bar{Z}}} \quad (44)$$

Here of course n_i is the number of reactive groups of class i , and k_i the corresponding intrinsic constant. The exponential factor, due to electrostatic interactions, is the same for every term in the sum. As with equation (12.1) it is apparent that the limit of $\bar{\nu}e^{-2w\bar{Z}}/(A)$, as $\bar{\nu} \rightarrow 0$, is equal to $\sum n_i k_i$, and the limit of $\bar{\nu}$, as $\bar{\nu}e^{-2w\bar{Z}}/(A) \rightarrow 0$, is equal to the total number of combining sites $n = \sum n_i$. Because of the curvature of the plot in Fig. 16, the extrapolated intercepts are subject to considerable uncertainty. The curve as drawn gives $\sum n_i k_i \cong 10,750$, and $n \cong 40$. Further analysis indicates that the curve can be fitted, within the moderately large limits of experimental uncertainty, by assuming only two classes of groups: class 1, with $n_1 = 10$ groups, and $k_1 = 1000$; and class 2, with

$n_2 = 30$ groups, and $k_2 = 25$. (The concentration units employed were moles per kilogram of H_2O , so k_1 and k_2 have the dimensions kilograms of H_2O per mole.)

The corresponding plot for chloride binding, from the data of Fig. 12, is not shown here. Scatchard *et al.* concluded that the data were best fitted by assuming the same two classes of groups ($n_1 = 10$ and $n_2 = 30$) as for thiocyanate binding, but with $k_1 = 44$ and $k_2 = 1.1$. The assumed

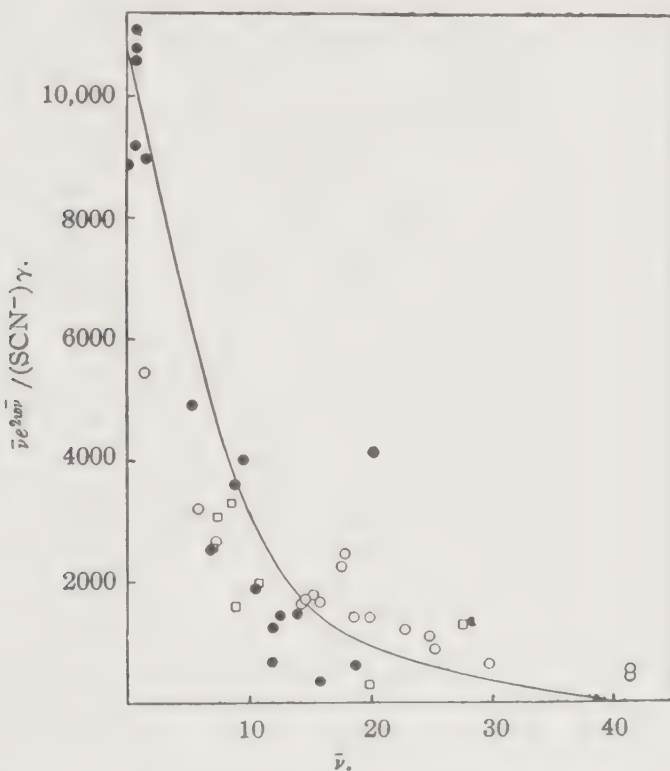


FIG. 16. The binding of thiocyanate ions by human serum albumin, plotted by the method of Fig. 6 and equation (16), after correction for the electrostatic interaction factor (equation 43). (From Scatchard, Scheinberg, and Armstrong, 1950.)

ratio $k_1/k_2 = 40$ is the same as for thiocyanate, but the absolute values of k_1 and k_2 are about twenty-three times as large for thiocyanate as for chloride. Actually the chloride data for isoionic albumin could be fitted equally well by assuming only one class of 11 groups, taking $k_1 = 44$; but the effects of variation in total charge on the albumin on the binding of chloride ion indicated that there were many more than 11 sites capable of binding chloride. This evidence is briefly discussed below.

The nature of the reactive groups on the albumin molecule is uncertain. Since the ions bound are anions, it is natural to suspect that positively charged groups on the albumin molecule are involved. There is no

obvious correlation, however, between the number of histidine, arginine, and lysine residues in the albumin molecule (see the discussion of the titration curve of albumin in Chapter 9) and the numbers of the groups with characteristic binding constants for anions. Furthermore, other proteins like the serum γ -globulins also contain many positively charged residues, yet have little or none of the affinity for anions that is displayed by serum albumins.

The binding of anions by albumin or other proteins is of course modified by the effects of binding of other ions on the net charge, \bar{Z} , as may be seen by considering equations (42) and (44). If \bar{Z} becomes more positive, owing for instance to binding of protons, the value of $\bar{\nu}$ for the anion, A, at constant (A), increases. Thus Scatchard *et al.* studied serum albumin at pH 3.2, where the bound hydrogen ions gave a value of $\bar{\nu}_H$ of 78, taking $\bar{\nu}_H = 0$ for isoionic albumin. Since they showed that there was no indication of the binding of sodium ions, the total net charge was $\bar{Z} = \bar{\nu}_H - \bar{\nu}_{Cl} = 78 - \bar{\nu}_{Cl}$. The chloride ion concentration was 0.1494 M. The measured value of $\bar{\nu}_{Cl}$ was 31, which agreed closely with the value calculated from (44), with $n_1 = 10$, $n_2 = 30$, and the k values given above.⁹ In contrast, the value of $\bar{\nu}_{Cl}$ for isoionic albumin at the same chloride ion concentration was near 7.5.

If \bar{Z} for serum albumin is negative, the value of $\bar{\nu}_{Cl}$ at a given value of (A) will obviously be less than for the isoionic protein. The binding of such anions as chloride by anionic albumin is not negligible, however. Under physiological conditions, in blood plasma, at pH 7.4, chloride ion concentration is near 0.10 M, w is near 0.03, and $\bar{\nu}_H$ is of the order of -20 . It then follows from (44), with the given values of the n 's and k 's, that $\bar{\nu}_{Cl}$ is of the order of 6 to 7; hence $\bar{Z} = -26$ to -27 . The weight concentration of albumin in human blood plasma is near 40 g/l; hence the molar concentration is near 6×10^{-4} , and the molar concentration of bound chloride is of the order of 0.004, or near 3% of the total chloride. This is by no means a negligible quantity in calculations of the distribution of chloride ion between plasma and red cells, or between plasma and other body fluids.

Recently Scatchard, Coleman, and Shen (1957) have studied the binding of several anions to bovine serum mercaptalbumin, making use of electromotive force measurements on ion exchanger electrode systems. Such ion exchangers are very useful in electrochemical studies on protein solutions, since the pores in the exchanger membranes are too fine to

⁹ Obviously, in this system, to solve equation (42) or (44) for $\bar{\nu}_{Cl}$ with known values of n_i , k_i , and w , it is necessary to proceed by trial and error. The value of $\bar{\nu}_{Cl}$ is a function of \bar{Z} , but \bar{Z} also depends on $\bar{\nu}_{Cl}$. Successive approximations lead rapidly to the correct result, however.

admit protein molecules, and they repel ions of the same sign as the charge on the membrane. Thus "even in the presence of protein, a cation exchanger membrane behaves as a small-cation electrode and an anion exchanger membrane as a small-anion electrode in the same sense that a glass membrane may behave as a hydrogen electrode" (Scatchard, Coleman, and Shen, p. 12). In these studies the isoionic bovine mercaptalbumin, which had been prepared by H. M. Dintzis, had been passed over a mixed-bed ion exchange resin, so that it was almost completely free of small ions, except for H^+ and OH^- ions. Moreover the greater part of the fatty acid, which generally remains quite tightly bound to serum albumin, had been removed by this treatment. This was probably the greatest difference between this albumin preparation and the human serum albumin studied by Scatchard, Scheinberg, and Armstrong (1950). The binding of ions by the bovine mercaptalbumin up to $\bar{\nu} = 27$ was described in terms of three classes of binding sites, with $n_1 = 1$, $n_2 = 8$, and $n_3 = 18$ sites per albumin molecule. For chloride ions the corresponding intrinsic binding constants were $k_1 = 2400$, $k_2 = 100$, and $k_3 = 3.3$. The same set of three classes of sites accounted for the binding of iodide, thiocyanate, and trichloroacetate, the ratio $k_1:k_2:k_3$ being the same in all cases. Each of the k values was 3.85 times as great for I^- as for Cl^- and 19.25 times as great for SCN^- and CCl_3COO^- as for Cl^- . Evidence for the presence of many other binding sites, with weaker affinities for anions, was also clearly shown in these studies, which were extended to solutions of high and low pH, in order to study the effect of varying the net charge on the protein over a wide range. The one site with a very high affinity for anions (i.e., the site with $k_1 = 2400$ for Cl^-) was probably made available for binding anions as a result of removal of fatty acid anions in the preliminary deionization process. Apart from this, the data of Scatchard, Coleman, and Shen for $\bar{\nu}$ as a function of the concentration of free anion in solution gave curves not very different from those of Scatchard, Scheinberg, and Armstrong, up to $\bar{\nu} = 12$ (see also Scatchard, 1955).

Effects of Competition between Different Ligands for the Same Binding Site

Frequently two or more substances are present in solution, both of which can attach themselves to the same binding sites on a macromolecule, P. The simplest such case occurs when all the n sites are equivalent and independent. If we call the two reacting species A and B, and the corresponding association constants k_A and k_B , then the binding of A in the presence of B is given by the equation

$$\bar{\nu}_A = \frac{nk_A(A)}{1 + k_A(A) + k_B(B)} \quad (45)$$

This can also be written

$$\bar{\nu}_A = \frac{nk_A'(A)}{1 + k_A'(A)} \quad (45.1)$$

where k_A' is given by

$$k_A' = \frac{k_A}{1 + k_B(B)} \quad (45.2)$$

The proof of these relations (compare equation 5) is straightforward and is left to the reader. Thus, in the presence of a constant concentration of B, the binding of A in such a system can be represented by equations of exactly the same form as (12), (15), or (16), except that the value of k_A' will be lower than that of the true constant, k_A , and is a function of (B). If k_A is known, and k_A' determined for a known value of (B), k_B can be determined from an equation obtained by rearranging (45.2):

$$k_B = \frac{1}{(B)} \left(\frac{k_A}{k_A'} - 1 \right) \quad (45.3)$$

If k_A' is determined at several different values of (B), and a constant value of k_B is then derived from (45.3), this gives a strong indication of the validity of the assumption that A and B are both reacting reversibly with the same set of sites on P. Thus Klotz *et al.* (1948) studied the effects of salicylate, dodecylsulfate, and other anions on the binding of methyl orange by serum albumin, and showed that relations equivalent to (45.3) held in these systems. They found k_A for methyl orange to be 4200 at pH 7.6 and 24°, and estimated n , the average number of combining sites per albumin molecule, as 12.8. The decrease in the binding of methyl orange, produced by the added competitor, could be followed by the change in absorption spectrum of the dye which accompanied binding. Thus they determined k_B for salicylate as 250 and for dedecyl sulfate as 2000. For the anion of acetyltryptophan the value found was 140, and for that of acetyl-leucine it was only 46.

Obviously an equation of the same form as (45) for the binding of B in the presence of A can be written:

$$\bar{\nu}_B = \frac{nk_B(B)}{1 + k_A(A) + k_B(B)} \quad (46)$$

Other relations, corresponding to (45.1), (45.2), and (45.3), follow from (46). From (45) and (46) it is obvious that the relation of bound A to bound B is

$$\frac{\bar{\nu}_A}{\bar{\nu}_B} = \frac{k_A(A)}{k_B(B)} \quad (47)$$

The extension of these relations to binding by molecules with several different classes of binding sites, or to systems in which electrostatic interactions are important, involves no new principles beyond those already set forth in this chapter. We may note, however, that, in systems containing buffers, one or both of the constituents of the buffer may act as competitors to the ligand which is under study. This is particularly true when binding by proteins such as serum albumin is being studied. The apparent affinity constants of such a protein for a given ligand, when determined in the presence of a buffer, are very likely to be lower than they would be if the buffer were not present. The effect of varying the total concentration of buffer, at constant pH , should always be studied in such cases if it is desired to know the true affinity constants. It may be possible to eliminate the use of buffers entirely, as in the work of Scatchard, Scheinberg, and Armstrong, which has been discussed above.

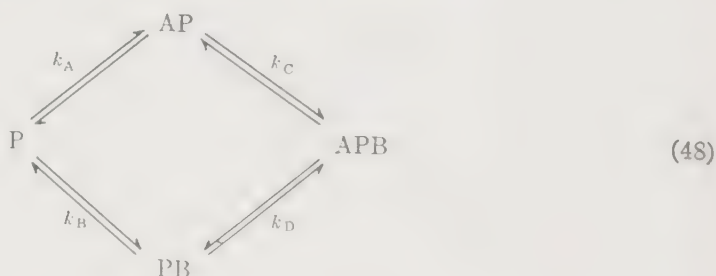
Linked Functions

Whenever a molecule possesses two or more different functions, e.g., dissociation of protons and combination with oxygen, belonging to nearby groups in the molecule, there is the likelihood of an interdependence of the functions due to interaction between the groups. When this occurs, we may conveniently speak of the functions as linked and refer to the groups as linked groups. Thus, in the example just given, the acid group may be referred to as an oxygen-linked acid group, and the oxygen-combining group, correspondingly, as an acid-linked or proton-linked group. Such groups in hemoglobin are accountable for the well-known Bohr effect, which will be discussed later. To take another example from hemoglobin, there is also a pH dependence of the reaction with fluoride, which leads us to speak of fluoride-linked acid groups and acid-linked fluoride-binding groups. Likewise we may speak of oxidation-linked acid groups and acid-linked oxidizable groups. The extreme case of linkage is that in which the two reactions are mutually exclusive—competing reactions—so that combination of the molecule with one reagent involves displacement of the other. Such a condition may be expected when both reactions involve one and the same group, as do the reactions of hemoglobin with carbon monoxide and oxygen. We might speak of such reactions as identically linked.

Clearly a study of the linkage among various reactions in a complex molecule possessing a variety of functions can be instructive in many ways as to the structure and characteristics of the molecule. This has proved to be particularly true in the case of the heme proteins, notably hemoglobin, which are exceptional in their capacity to take part in diverse reactions. When one of the functions involves a large number of

groups, the study of linkage often serves to single out certain of these groups from the rest, namely, those which are linked with some other function. It provides, as it were, a technique for spotlighting a special class of the groups in question. For example, if we are interested primarily in the acid groups, then by studying the linkage of proton dissociation with a variety of other functions, e.g., oxidation, combination with oxygen, carbon monoxide, and fluoride, over a wide pH range, we may reveal selectively the existence of a variety of acid groups, each related to one or more of the other functions. Often, on the basis of other considerations and special argument, it is possible to identify these acid groups with particular amino acid residues in the protein and to show how they are affected by reactions involving other groups. This principle forms the basis of the so-called differential titration, which has contributed so much to an understanding of the heme proteins. Since linkage often implies proximity, the phenomenon may provide suggestive information on the special relationship of the different groups in the molecule, or on changes of configuration which accompany the reaction process. The study of other kinds of linkage than that involving proton dissociation has also proved instructive.

The simplest case of linkage to be considered is that of a molecule, P , containing just two reactive groups, one of which can combine with a ligand A , the other with a ligand B . We denote the compound of P with A by AP , that with B by PB , and that with both by APB . Then the interactions involved may be represented by the scheme



$$k_A = \frac{(AP)}{(P)(A)}; k_B = \frac{(PB)}{(P)(B)}; k_C = \frac{(APB)}{(AP)(B)}; k_D = \frac{(APB)}{(PB)(A)} \quad (49)$$

Here it is obvious that

$$k_A k_C = k_B k_D = \frac{(APB)}{(P)(A)(B)} \quad (49.1)$$

Equation (49) directly implies an important reciprocal relation. If, for instance, $k_D > k_A$, this means that the attachment of B to P increases the affinity of P for A . But if this relation holds, it follows from

(49) that $k_C > k_B$, since $k_C/k_B = k_D/k_A$. Hence if the binding of B increases the affinity of P for A, the binding of A increases the affinity of P for B. Either relation implies the other.

It is not necessary, of course, that any detectable interaction should occur in such a system. The binding of A may be totally independent of that of B. Equation (49) merely requires that, *if* the binding of A affects that of B, then the reciprocal relation must also hold.

Now consider the quantity $\bar{\nu}_A$ —that is, the amount of A bound per molecule of P present in the system in all forms—namely P, AP, PB, and APB. Since P is monovalent for A, $\bar{\nu}_A$ can vary only between zero and 1.

$$\begin{aligned}\bar{\nu}_A &= \frac{(\text{AP}) + (\text{APB})}{(\text{P}) + (\text{AP}) + (\text{PB}) + (\text{APB})} \\ &= \frac{k_A(\text{A}) + k_A k_C(\text{A})(\text{B})}{1 + k_A(\text{A}) + k_B(\text{B}) + k_A k_C(\text{A})(\text{B})} \quad (50)\end{aligned}$$

Now consider the value of $\bar{\nu}_A$, in the absence of B, at a given activity of free A in solution. Setting $(\text{B}) = 0$ in (49), we obtain the familiar equation

$$(\bar{\nu}_A)_{\text{B}=0} = \frac{k_A(\text{A})}{1 + k_A(\text{A})} \quad (51)$$

On the other hand, consider the value of $\bar{\nu}_A$ if (B) becomes so large that all molecules of P are saturated with B. In that case we let (B) in equation (50) become infinite; or we note in the reaction scheme (48) that the only species of P that can exist at appreciable concentration under these conditions are PB and APB. It follows from either of these considerations that

$$(\bar{\nu}_A)_{\text{B} \rightarrow \infty} = \frac{k_D(\text{A})}{1 + k_D(\text{A})} \quad (52)$$

Thus, if we measure $\bar{\nu}_A$ at a given value of (A) , first in the absence of B and then when P is saturated with B, the increment in $\bar{\nu}_A$ is

$$\Delta \bar{\nu}_A = \frac{k_D(\text{A})}{1 + k_D(\text{A})} - \frac{k_A(\text{A})}{1 + k_A(\text{A})} \quad (53)$$

This is the essence of a differential titration. If A is hydrogen ion and B is oxygen, for example, P in the hypothetical system just discussed would be a sort of simplified hemoglobin molecule, with only one heme group and one heme-linked acid group. The value of $\Delta \bar{\nu}_A$ would then measure the number of protons bound (or released) per molecule of P, when P passes from the completely deoxygenated to the completely oxygenated state.

Problem. Take $k_A = 10^8$ and $k_D = 10^7$; compute $\Delta\bar{\nu}_H$ from (53) for various values of (H^+) , and plot $\Delta\bar{\nu}_H$ as a function of pH . Repeat for the case $k_A = 10^8$, $k_D = 10^6$. At what value of (A) is $\Delta\bar{\nu}_A$ a maximum (or minimum)? What is the maximum (or minimum) value of $\Delta\bar{\nu}_A$?

The value of $\bar{\nu}_B$ is given by an equation corresponding to (50):

$$\bar{\nu}_B = \frac{(PB) + (APB)}{(P) + (AP) + (PB) + (APB)} = \frac{k_B(B) + k_A k_C(A)(B)}{1 + k_A(A) + k_B(B) + k_A k_C(A)(B)} \quad (54)$$

Hence the increment in $\bar{\nu}_B$, at a given value of (B) , from a system in which $(A) = 0$ to a system in which P is saturated with A , is

$$\Delta\bar{\nu}_B = (\bar{\nu}_B)_{A=\infty} - (\bar{\nu}_B)_{A=0} = \frac{k_C(B)}{1 + k_C(B)} - \frac{k_B(B)}{1 + k_B(B)} \quad (55)$$

The competition of two ligands for the same site on P may be treated as a special case of scheme (48). In this case either A or B can combine with P , but not both; hence APB cannot exist, and $k_C = k_D = 0$. Then (50) becomes

$$\bar{\nu}_A = \frac{k_A(A)}{1 + k_A(A) + k_B(B)} \quad (56)$$

with a corresponding equation for $\bar{\nu}_B$. These equations are identical with equations (45) and (46), which have already been obtained for a system involving competition between A and B , for the case $n = 1$.

We now consider the more general case of two linked functions, if the molecule P contains n sites at which A can combine and r sites at which B can combine. The various molecular species which may exist in a solution containing P , A , and B may be denoted by the formula A_iPB_j , where i may have any value from zero to n , and j any value from zero to r . Moreover, for specified values of i and j , the molecules of the class denoted by A_iPB_j represent the total of a large number of microscopically different forms. As we have seen in the discussion in Chapter 9, we may distribute i indistinguishable A molecules among n sites in $n!/i!(n-i)!$ different ways, and j B molecules among r sites in $r!/j!(r-j)!$ different ways. Since any distribution in the former set may be combined with any of the latter, the total number of microscopic forms (w_{ij}) corresponding to the general formula for the class A_iPB_j is

$$w_{ij} = \frac{n!}{i!(n-i)!} \frac{r!}{j!(r-j)!} \quad (57)$$

The total number of microscopic forms of all classes, from P to A_nPB_r , is 2^{n+r} .

Problem. Suppose there is only one class of n sites, each of which can bind either an A molecule or a B molecule, but not both. Show that in this case the number of microscopic forms corresponding to the formula A_iPB_j is

$$w_{ij} = \frac{n!}{i!j!(n-i-j)!}; \quad (i+j) \leq n \quad (58)$$

It is not necessary for our present purposes, however, to consider all these different microscopic forms in calculating the binding of A and B. We may formulate the binding of A by taking the total concentration of all molecules of the class A_iPB_j , multiplying by i , the number of A molecules bound by molecules of this class, and adding these terms for all classes. This sum, divided by the total concentration of P molecules in every form from P to A_nPB_r , gives $\bar{\nu}_A$. A similar summation, in which each term in the numerator is multiplied by j , gives $\bar{\nu}_B$.

We define a set of affinity constants, L_{ij}^* , by a natural extension of the definition of the constants L_i^* given in equation (24):

$$L_{ij}^* = \frac{(A_iPB_j)}{(A)^i(B)^j(P)} \quad (59)$$

For the binding of A we may then write

$$\bar{\nu}_A = \sum_{i=1}^n \sum_{j=0}^r iL_{ij}^*(A)^i(B)^j / \sum_{i=0}^n \sum_{j=0}^r L_{ij}^*(A)^i(B)^j = \frac{\partial \ln S}{\partial \ln (A)} \quad (60)$$

Here S denotes the double sum in the denominator. In taking the partial derivative, (B) is held constant. Also, of course, there is an exactly analogous relation involving B:

$$\bar{\nu}_B = \frac{\partial \ln S}{\partial \ln (B)} \quad (61)$$

the differentiation being carried out at constant (A).

Since S is determined completely by (A) and (B), we obtain by cross differentiation the relation

$$\frac{\partial^2 \ln S}{\partial \ln (A) \cdot \partial \ln (B)} = \frac{\partial \bar{\nu}_B}{\partial \ln (A)} = \frac{\partial \bar{\nu}_A}{\partial \ln (B)} \quad (62)$$

This shows in its most general form the reciprocal relation between the two linked functions. For instance, equation (62) indicates immediately that, if increasing the oxygen activity causes release of protons from a hemoglobin molecule at fixed (H^+) , then increase of (H^+) —that is, decrease of pH —causes oxygen to be released from oxyhemoglobin at a fixed oxygen activity. Since of the variables (A), (B), $\bar{\nu}_A$, and $\bar{\nu}_B$ only two

are independent, this relation may be thrown into a variety of other forms. One useful one is

$$-\left(\frac{\partial \bar{v}_A}{\partial \bar{v}_B}\right)_{(A)} = \left[\frac{\partial \ln (B)}{\partial \ln (A)}\right]_{\bar{v}_B} \quad (63)$$

In its application to hemoglobin, this shows, for instance, how the effect of oxygenation on the dissociation of protons may be expressed in terms of the change with pH of the logarithm of the oxygen pressure required to maintain a given degree of saturation of the hemoglobin with oxygen.

Another way of expressing the relation between linked functions is often useful. Consider the molecules of P which are combined with just j molecules of B , regardless of the amount of A combined, and denote their concentration by (ΣPB_j) , which is an abbreviated notation for

$$\sum_{i=0}^n (A_i PB_j):$$

$$\begin{aligned} (\Sigma PB_j) &= (PB_j) + (APB_j) + (A_2PB_j) + \cdots \\ &\quad + (A_nPB_j) \equiv \sum_{i=0}^n (A_i PB_j) \end{aligned} \quad (64)$$

This may be written, making use of (59), as

$$(\Sigma PB_j) = (P)(B)^j \sum_{i=0}^n L_{ij}^*(A)^i \quad (65)$$

Here (P) of course denotes the class of molecules which are uncombined with either A or B . This expression is valid for all values of j , including zero. Therefore we may write

$$\frac{(\Sigma PB_j)}{(\Sigma PB_0)(B)^j} = \frac{\sum_{i=0}^n L_{ij}^*(A)^i}{\sum_{i=0}^n L_{j0}^*(A)^i} \equiv R_j \quad (66)$$

It will be seen that the right-hand member of this equation, which we have denoted by R_j , is the apparent constant for the equilibrium



From equation (27) it will be seen that, formally at least, we may write for the numerator of the expression for R_j an expansion in terms of a series of titration constants, $g_{A_1} \dots g_{A_n}$:

$$\Sigma L_{ij}^*(A)^i = L_{0j}^*[1 + g_{A_1}(A)][1 + g_{A_2}(A)] \dots [1 + g_{A_n}(A)] \quad (67)$$

Here L_{0j}^* is simply the constant for the restricted equilibrium involving the molecules of uncombined P and molecules of the class A_0PB_j , which contain j molecules of bound B, and no bound A. It is independent of (A) , in so far as activity coefficients vary with (A) . The denominator of R_j in (66) is given by a similar expression in which L_{0j}^* is replaced by unity ($L_{00}^* = 1$). The equilibrium constant, R_j , may therefore be written

$$R_j = L_{0j}^* \frac{\{[1 + g_{A_1}(A)] \dots [1 + g_{A_n}(A)]\}(\Sigma PB_j)}{\{[1 + g_{A_1}(A)] \dots [1 + g_{A_n}(A)]\}(\Sigma PB_0)} \quad (68)$$

The g 's in the numerator will in general be different from those in the denominator, since the former refer to all members of the class

$\sum_{i=0}^n A_iPB_j$, whereas the latter refer to all members of the class $\sum_{i=0}^n A_iPB_0$.

Equation (68) brings out clearly the dependence of the constants for the equilibrium involving B on those involving A.

Equations (67) and (68) are subject to the same restrictions already pointed out in discussing (25)—namely, that the titration constants g_{A_1} , g_{A_2} , etc., are real positive numbers only if they form a series in descending order of magnitude: $g_A \geq g_{A_2} \geq g_{A_3} \dots$. It is only when this relation holds that (67) and (68) are likely to be of much use in practice: otherwise we must employ (66) for R_j , since it is not subject to any such restriction. When (68) is applicable, it may have considerable advantages over (66) in formulating the relations involved. For instance many of the titration constants describing the interaction of P with A may be unaffected by combination of P with B. If this is true of the i th titration constant, then g_{A_i} is the same in the numerator and denominator of (68), and the corresponding terms may be canceled out of the equation. Most of the acidic groups in oxyhemoglobin and reduced hemoglobin, for instance, have the same g values; we can eliminate them from consideration, and consider only the heme-linked groups, when we consider the effect of pH on the affinity constants for oxygen.

When (68) is applicable, it may be thrown into another form by use of (60):

$$\frac{\partial \ln R_j}{\partial \ln (A)} = (\bar{\nu}_A)_j - (\bar{\nu}_A)_0 = \sum_{i=1}^n \left[\frac{g_{A_i}(A)}{1 + g_{A_i}(A)} \right]_{\Sigma PB_j} - \sum_{i=1}^n \left[\frac{g_{A_i}(A)}{1 + g_{A_i}(A)} \right]_{\Sigma PB_0} \quad (69)$$

This expresses the variation of the equilibrium constant, R_j with (A) in terms of $(\bar{\nu}_A)_j$, the amount of A combined with molecules of the class ΣPB_j , and $(\bar{\nu}_A)_0$, the amount of A combined with molecules of the class PB_0 . In subsequent chapters these relations will be used to formulate the effect of pH on equilibria involving oxidation-reduction processes and others which involve the mutual interactions of hydrogen ions and oxygen—or carbon monoxide—on hemoglobin.

GENERAL REFERENCES

- Bailar, J. C., Jr. (ed.) (1956). "The Chemistry of the Coordination Compounds," Reinhold Publishing Co., New York. A large, comprehensive monograph. Several chapters will be found very useful in connection with the subjects discussed here.
- Gurd, F. R. N. (ed.) (1954). "Chemical Specificity in Biological Interactions," Academic Press, New York. This book contains several valuable articles, especially those by C. D. Coryell, J. Schubert, G. Schwarzenbach, and the final chapter by G. Scatchard, W. L. Hughes, Jr., F. R. N. Gurd, and P. E. Wilcox. A comprehensive and critical review of this important field.
- Gurd, F. R. N., and Wilcox, P. E. (1956). Complex formation between metallic cations and proteins, peptides and amino acids, *Advances in Protein Chem.* **11**, 311.
- Klotz, I. M. (1953). Protein interactions, in "The Proteins" (H. Neurath and K. Bailey, eds.), Vol. IB, Chapter 8, Academic Press, New York.
- Martell, A. E., and Calvin, M. (1952). "Chemistry of the Metal Chelate Compounds," Prentice-Hall, New York. A valuable and extensive review of this important field.
- Sørensen, S. P. L., Linderstrøm-Lang, K., and Lund, E. (1926). *Compt. rend. trav. lab. Carlsberg (Sér. chim.)* **16**, No. 5.
- Sørensen, S. P. L., Linderstrøm-Lang, K., and Lund, E. (1927). *J. Gen. Physiol.* **8**, 543.
- Waugh, D. F. (1954). Protein-protein interactions, *Advances in Protein Chem.* **9**, 325.

OTHER REFERENCES

- Alberty, R. A. (1953). Electrochemical properties of the proteins and amino acids, in "The Proteins" (H. Neurath and K. Bailey, eds.), Vol. IA, Chapter 6, Academic Press, New York.
- Bjerrum, J. (1941). "Metal Ammine Formation in Aqueous Solution," P. Haase and Son, Copenhagen.
- Bjerrum, J. (1950). *Chem. Revs.* **46**, 381.
- Boyd, W. C. (1956). "Fundamentals of Immunology," 3rd ed., Interscience Publishers, New York.
- Chanutin, A., Ludewig, S., and Masket, A. V. (1942). *J. Biol. Chem.* **143**, 737.

- Datta, S. P., and Rabin, B. R. (1956). *Trans. Faraday Soc.* **52**, 1117, 1123; *Biochim. et Biophys. Acta* **19**, 572.
- Dobbie, H., and Kermack, W. O. (1955). *Biochem. J.* **59**, 246.
- Eadie, G. S. (1942). *J. Biol. Chem.* **146**, 85.
- Edelhoch, H., Katchalski, E., Maybury, R. H., Hughes, W. L., Jr., and Edsall, J. T. (1953). *J. Am. Chem. Soc.* **75**, 5058.
- Edsall, J. T. (1954). Interaction of proteins and ions with special reference to mercury derivatives of mercaptalbumin, in "Ion Transport across Membranes" (H. T. Clarke, ed.), p. 221, Academic Press, New York.
- Edsall, J. T., Felsenfeld, G., Goodman, D. S., and Gurd, F. R. N. (1954a). *J. Am. Chem. Soc.* **76**, 3054.
- Edsall, J. T., Maybury, R. H., Simpson, R. B., and Straessle, R. (1954b). *J. Am. Chem. Soc.* **76**, 3131.
- Gurd, F. R. N., and Goodman, D. S. (1952). *J. Am. Chem. Soc.* **74**, 670.
- Gurd, F. R. N., and Murray, G. R. (1954). *J. Am. Chem. Soc.* **76**, 187.
- Haurowitz, F. (1951). *J. Biol. Chem.* **193**, 443.
- Hayes, J. E., Jr., and Velick, S. F. (1954). *J. Biol. Chem.* **207**, 225.
- Hearon, J. Z., and Gilbert, J. B. (1955). *J. Am. Chem. Soc.* **77**, 2594.
- Hughes, W. L., Jr. (1947). *J. Am. Chem. Soc.* **69**, 1836.
- Hughes, W. L., Jr. (1950). *Cold Spring Harbor Symposia Quant.* **14**, 79.
- Kay, C. M., and Edsall, J. T. (1956). *Arch. Biochem. and Biophys.* **65**, 354.
- Klotz, I. M., Triwush, H., and Walker, F. M. (1948). *J. Am. Chem. Soc.* **70**, 2935.
- Longworth, L. G., and Jacobsen, C. F. (1949). *J. Phys. & Colloid Chem.* **53**, 126.
- Michaelis, L. (1946). In "Currents in Biochemical Research" (D. E. Green, ed.), Chapter 14, Interscience Publishers, New York.
- Nozaki, Y., Gurd, F. R. N., Chen, R., and Edsall, J. T. (1957). *J. Am. Chem. Soc.* **79**, 2123.
- Putnam, F. W. (1953). Protein denaturation, in "The Proteins" (H. Neurath and K. Bailey, eds.), Vol. IB, Chapter 9, Academic Press, New York.
- Rabin, B. R. (1956). *Trans. Faraday Soc.* **52**, 1130.
- Scatchard, G. (1949). *Ann. N. Y. Acad. Sci.* **51**, 660.
- Scatchard, G. (1955). Transport of ions across charged membranes, in "Electrochemistry in Biology and Medicine" (T. Shedlovsky, ed.), p. 18, John Wiley & Sons, New York.
- Scatchard, G., and Black, E. S. (1949). *J. Phys. & Colloid Chem.* **53**, 88.
- Scatchard, G., Coleman, J. S., and Shen, A. L. (1957). *J. Am. Chem. Soc.* **79**, 12.
- Scatchard, G., Hughes, W. L., Jr., Gurd, F. R. N., and Wilcox, P. E. (1954). The interaction of proteins with small molecules and ions, in "Chemical Specificity in Biological Interactions" (F. R. N. Gurd, ed.), Chapter XI, Academic Press, New York.
- Scatchard, G., Scheinberg, I. H., and Armstrong, S. H. (1950). *J. Am. Chem. Soc.* **72**, 535, 540.
- Schwarzenbach, G. (1954). The specificity of metal complex formation, in "Chemical Specificity in Biological Interactions" (F. R. N. Gurd, ed.), Chapter X, Academic Press, New York.
- Sen, D. N., Mizushima, S., Curran, C., and Quagliano, J. V. (1955). *J. Am. Chem. Soc.* **77**, 211.
- Sillén, L. G. (1949). *Acta Chem. Scand.* **3**, 539.
- Stricks, W., and Kolthoff, I. M. (1953). *J. Am. Chem. Soc.* **75**, 5673.
- Stricks, W., Kolthoff, I. M., and Heyndrickx, A. (1954). *J. Am. Chem. Soc.* **76**, 1515.

- Tanaka, N., Kolthoff, I. M., and Stricks, W. (1955). *J. Am. Chem. Soc.* **77**, 1996.
 Vallee, B. L., and Hoch, F. L. (1955). *Proc. Natl. Acad. Sci. (U.S.)* **41**, 327.
 Williams, R. J. P. (1953). *Biol. Revs. Cambridge Phil. Soc.* **28**, 381.
 Wyman, J. (1948). *Advances in Protein Chem.* **4**, 407.

SUPPLEMENTARY BIBLIOGRAPHY FOR THE DISCUSSION OF
 RIBONUCLEASE AND HEMOGLOBIN, PP. 97-99

- Allende, J. E., and Richards, F. M. (1962). *Biochemistry* **1**, 295
 Anfinsen, C. B. (1962). "Some Observations on the Basic Principles of Design in Protein Molecules," Symposium No. 4, Fifth International Conference of Biochemistry, Moscow, 1961. Pergamon, New York, in press.
 Braunitzer, G., Gehring-Muller, R., Hilschmann, N., Hilse, K., Hobom, G., Rudloff, J., and Wittmann-Liebold, B. (1961). *Z. physiol. Chem.* **325**, 283.
 Goldstein, J., Guidotti, G., Konigsberg, W., and Hill, R. J. (1961). *J. Biol. Chem.* **236**, PC77.
 Gross, E., and Witkop, B. (1962). *J. Biol. Chem.* **237**, 1856.
 Haber, E., and Anfinsen, C. B. (1962). *J. Biol. Chem.* **237**, in press.
 Hirs, C. H. W., Moore, S., and Stein, W. H. (1960). *J. Biol. Chem.* **235**, 633
 Konigsberg, W., Guidotti, G., and Hill, R. J. (1961). *J. Biol. Chem.* **236**, PC55.
 Potts, J. T., Berger, A., Cooke, J., and Anfinsen, C. B. (1962). *J. Biol. Chem.* **237**, 1851.
 Richards, F. M., and Vithayathil, P. J. (1959). *J. Biol. Chem.* **234**, 1459 and 2343; (1960) *Ibid.* **235**, 1029; (1961), *Ibid.* **236**, 1380; also (1960) *Brookhaven Symposia in Biol.* **13**, 115.
 Smyth, D. G., Stein, W. H., and Moore, S. (1962). *J. Biol. Chem.* **237**, 1845.
 Spackman, D. H., Stein, W. H., and Moore, S. (1960). *J. Biol. Chem.* **235**, 648.
 White, F. H., Jr. (1961). *J. Biol. Chem.* **236**, 1353.

Author Index

Numbers in parentheses are reference numbers and are included to assist in locating references when authors' names are not mentioned in the text. Numbers in italics refer to pages on which the references are listed.

A

Abelson, P. H., 15, 16, *25*
 Acree, S. F., 452(5), *453*
 Alberty, R. A., 213, 214, *239*, *240*, 399,
404, *467*, *475*, 599, *660*
 Albritton, E. C., *18*
 Alfrey, T., 542, *547*
 Allen, T. W., 174, *240*, 320, *322*
 Alpher, R. A., *25*
 Anfinsen, C. B., 83, 84, 86, 92, 93, 96, 97,
 98, 99, 128, 129, *133*, *135*, 530, *547*
 Anson, M. L., *132*
 Aoki, K., 536, *548*
 Armstrong, S. H., Jr., 519, 535, *548*, 597,
 598, 613, 646, 647, 649, 651, *661*
 Ashby, J. H., 452(6, 7), *453*
 Astbury, W. T., 101, 104, 107, 108, *133*

B

Bailar, J. C., Jr., *660*
 Bailey, K., *132*
 Baldwin, E., *25*
 Bamford, C. H., 119, *133*, 542, *547*
 Barnafi, L., 95, *133*
 Bascuas, G., *465*
 Bassham, J. A., 551, *589*
 Bates, R. G., 420, 441, 447, *448*, 452(2, 4,
 5, 8, 10, 11), *453*, 473, *475*
 Bear, R. S., 116, 117, 118, 119, 122, *133*
 Belcher, D., *558*
 Bell, P. H., 95, *133*, *134*
 Bell, R. P., *475*
 Benedict, M. A., 281, *322*
 Benesch, R., 88, *133*, *465*, 499, 500, 501,
502, *547*
 Benesch, R. E., 88, *133*, *465*, 499, 500
 501, 502, *547*
 Benzinger, T. H., 211, 212, *239*
 Berg, D., 559, *590*
 Berg, P. W., 542, *547*

Bernal, J. D., 14, 16, *25*, *45*
 Bjerrum, J., 626, 633, 634, 638, 641, *660*
 Bjerrum, N., 459, *476*
 Black, E. S., 600, 601, *661*
 Blanchard, M. H., *459*, 511, *548*, 566, *590*
 Block, R. J., 67, *133*
 Blum, H. F., 14, 16, *25*
 Bock, R. M., *239*, *467*, *475*
 Boedtker, H., 117, 123, *133*
 Böttcher, C. J. F., *385*
 Bonner, F. T., *453*, *555*
 Booth, V. H., 584, *590*
 Borsook, H., 209, 237, 238, *239*
 Bowen, J. M., 6, *25*
 Bower, V. E., *448*
 Boyd, W. C., 593, *660*
 Boyer, P. D., 215, *240*
 Bradbury, J. H., 129, *133*
 Bragg, W. L., *36*, 100, 132, *133*
 Brand, E., *504*
 Bricas, E., 78, *133*
 Bromer, W. W., 95, *133*
 Brown, H., 95, *133*
 Brown, H. C., *475*, 498, *547*
 Brown, L., 119, *133*
 Bunn, C. W., 100, *133*
 Burton, K., 237, 238, *239*
 Butler, J. A. V., 188, *191*
 Buzzell, J. G., 536, *549*

C

Calvin, M., *465*, 551, *589*, 609, 638, 639,
660
 Cannan, R. K., *465*, 516, 517, 522, 523,
 526, 527, 529, 530, 534, *547*
 Cant, E. M., 119, *133*
 Carr, C. W., 528, *547*
 Chance, B., 581, *590*
 Chanutin, A., 598, *660*
 Chen, R. F., *464*, *568*, 627, 631, 632, 634,
661

Clark, A. M., 587, 590
 Clark, W. M., 411, 420, 430, 475
 Clarke, H. B., 452(7), 453
 Cohen, C., 130, 131, 133, 135
 Cohn, E. J., 55, 132, 174, 197, 199, 239,
 274, 275, 276, 278, 295, 305, 306, 309,
 312, 314, 315, 317, 318, 319, 320, 321,
 372, 373, 377, 385, 407, 441, 458, 459,
 463, 464, 465, 466, 473, 475, 486, 488,
 509, 511, 529, 548, 566, 590
 Cole, R. D., 73, 95, 134, 135
 Coleman, J. S., 603, 619, 650, 661
 Conway, E. J., 19-22, 25
 Corey, R. B., 103, 105, 106, 109, 110, 134,
 135, 308, 313, 321
 Coryell, C. D., 640, 660
 Cowan, P. M., 118, 119, 122, 133
 Craig, L. C., 96, 134
 Crammer, J. L., 89, 133, 464, 467, 476,
 530, 548
 Crick, F. H. C., 101, 119, 120, 132, 133,
 135
 Crook, E. M., 452(6, 7), 453
 Curran, C., 643, 661

D

Datta, S. P., 452(6, 7), 453, 643, 661
 Davies, D. S., 95, 134
 Davis, R., 453, 555
 Davis, T. L., 465
 Dawes, E. A., 240
 Debye, P., 263, 264, 274, 278, 282ff.,
 321, 330, 336, 341, 385
 Denbigh, K. G., 228, 238
 Dintzis, H. M., 373, 600
 Dixon, J. S., 95, 134
 Dobbie, H., 643, 644, 661
 Dobry, A., 208, 240
 Donohue, J., 114, 133
 Doty, P., 117, 123, 129, 130, 133, 135
 Durrum, E. L., 67, 133
 du Vigneaud, V., 79, 133
 Dymond, J. A., 6, 25

E

Eadie, G. S., 622, 661
 Ebert, L., 486, 548
 Edelhoch, H., 645, 661

Edsall, J. T., 51, 55, 86, 111, 115, 125,
 127, 132, 133, 134, 174, 197, 199, 239,
 240, 274, 276, 295, 305, 306, 309, 312,
 314, 315, 317, 318, 319, 320, 321, 372,
 373, 377, 385, 407, 441, 458, 459, 463,
 464, 465, 466, 473, 475, 486, 488, 504,
 509, 511, 524, 529, 548, 568, 627, 631,
 632, 634, 645, 661
 Eigner, E. A., 95, 134
 Eisenberg, H., 542, 548
 Elderfield, R. C., 465
 Ellenbogen, E., 466
 Elliott, A., 119, 129, 133, 542, 547
 England, A., Jr., 278, 321
 Epprecht, A., 462, 476
 Epstein, J., 534, 540, 549
 Evans, E. A., Jr., 205, 240
 Everett, D. H., 452(9), 453

F

Failey, C. F., 274, 322
 Faurholt, C., 553, 571, 572, 575, 579, 580,
 590
 Felsenfeld, G., 627, 631, 634, 661
 Felsing, W. A., 453
 Ferguson, J. K. W., 571, 572, 577, 578, 590
 Fevold, H. L., 530, 548
 Field, E. O., 540, 548
 Fitts, D. D., 129, 133
 Fleming, H., 18, 26
 Flory, P. J., 231, 240, 374, 385
 Foster, J. F., 536, 548, 549
 Fowler, R. H., 45, 228, 239, 321
 Fraenkel-Conrat, H., 524, 548
 French, D., 524, 548
 French, D. M., 558, 559, 560
 Fröhlich, H., 385
 Fromageot, C., 78, 133, 464, 540, 548
 Fruton, J. S., 208, 240
 Fuoss, R. M., 542, 548

G

Gahatia, M., 542, 549
 Gane, R., 458
 Gawron, O., 465
 Geschwind, I. I., 95, 133, 134
 Gilbert, J. B., 630, 633, 661
 Glasstone, S., 174, 194, 228, 233, 288, 272,
 274, 321

Goldschmidt, V. M., 9, 26
 Goldwater, W. H., 504
 Goodman, D. S., 627, 631, 634, 642, 661
 Gorham, A. T., 452(14), 453
 Grafius, M. A., 498, 548
 Green, A. A., 51, 133, 277, 322, 511, 548, 566, 590
 Green, H., 215, 240
 Green, N. M., 75, 134
 Greenstein, J. P., 61, 134, 377, 426, 476
 Grenville-Wells, H. J., 112, 113, 114, 134
 Gross, J., 118, 135
 Gross, Ph., 271, 281, 322
 Gucker, F. T., Jr., 174, 240, 320, 322
 Guggenheim, E. A., 228, 238, 239, 321
 Gurd, F. R. N., 464, 568, 606, 627, 631, 632, 634, 638, 639, 640, 642, 643, 644, 660, 661
 Gustavson, K. H., 116, 117, 134

H

Häfliger, O., 475, 498, 547
 Hammett, L. P., 413, 473, 475
 Hanby, W. E., 119, 129, 133, 542, 547
 Harfenist, E. J., 73, 96, 134
 Harned, H. S., 291, 322, 405, 433, 441, 448, 452(1), 453, 472, 473, 475, 555, 561
 Harrington, W. F., 125, 134
 Harris, F. E., 542, 548
 Harris, J. I., 95, 134
 Hastings, A. B., 558, 584, 585, 590
 Hauenstein, J. D., 534, 536, 537, 549
 Haurowitz, F., 605, 661
 Hayes, J. E., Jr., 598, 619, 620, 661
 Hearon, J. Z., 630, 633, 661
 Hems, R., 212, 239
 Henderson, L. J., 23, 26, 37, 41, 45, 540, 548, 555, 556, 571, 589
 Herman, R. C., 25
 Hermans, J. J., 542, 548
 Heyndrickx, A., 644, 662
 Highberger, J. H., 118, 135
 Hildebrand, J. H., 189, 240
 Hill, T. L., 495, 522, 548
 Hirs, C. H. W., 53, 70, 73, 92, 97, 98, 134
 Hoch, F. L., 620, 662
 Hollingworth, B. R., 504
 Holtzer, A. M., 129, 133

Howard, K. S., 95, 134
 Howells, E. R., 132, 133
 Hückel, E., 282ff., 321
 Hughes, W. L., Jr., 51, 63, 133, 134, 532, 548, 606, 645, 660, 661
 Hutchinson, G. E., 10, 25, 26
 Hvidt, A., 127, 128, 134

I

Ingold, C. K., 458, 462, 463, 475
 Ingram, V. M., 88, 99, 134
 Iser, M., 271, 272, 281, 322

J

Jacobsen, C. F., 599, 661
 Jeans, J. H., 321
 Johnson, W., 18, 26
 Jones, A. V., 452(3), 453
 Jordan, D. O., 467
 Joseph, N. R., 314, 322

K

Kartha, G., 118, 135
 Katchalski, E., 542, 543, 544, 548, 549, 645, 661
 Katchalsky, A., 542, 545, 548
 Katsoyannis, P. G., 79, 133
 Kauzmann, W., 89, 134
 Kay, C. M., 645, 661
 Kendrew, J. C., 101, 104, 105, 107, 110, 116, 118, 133, 134, 610
 Kermack, W. O., 643, 644, 661
 Kibrick, A. C., 516, 517, 522, 523, 526, 527, 529, 530, 534, 547
 Kiese, M., 583, 584, 585, 590
 King, E. J., 452(12), 453, 476
 King, G. W., 476
 Kirby, A. H. M., 464
 Kirkwood, J. G., 129, 133, 305, 314, 315, 316, 322, 356, 361, 385, 461, 476, 546, 549
 Kitai, R., 95, 96, 133, 135
 Klemperer, F. W., 377
 Klotz, I. M., 174, 239, 240, 320, 322, 488, 548, 596, 652, 660, 661
 Knight, B. C. J. G., 465
 Kolthoff, I. M., 644, 645, 661

Kornberg, H. L., 237, 238, 239, 552, 589
 Krebs, H. A., 205, 206, 207, 237, 238, 239,
 240, 552, 589
 Kunkel, H. G., 73, 135

L

Lardy, H. A., 88, 133
 Larsson, E., 464, 465
 Laskowski, M., Jr., 510, 545, 548
 Lederer, E., 67, 134
 Lederer, M., 67, 134
 Levene, P. A., 467
 Levintow, L., 211, 213, 240
 Levy, A. L., 67, 134
 Lewin, J., 320, 322
 Lewis, G. N., 239
 Li, C. H., 95, 96, 133, 134
 Li, N. C., 465
 Libby, W. F., 6, 26
 Lifson, S., 542, 548
 Linderstrøm-Lang, K., 127, 134, 305, 322,
 515, 548, 605, 660
 Lindley, H., 126, 134
 Loeb, G. I., 536, 547, 548
 Long, F. A., 282, 322
 Longworth, L. G., 599, 661
 Lonsdale, K., 100, 134
 Low, B. W., 86, 101, 108, 110, 111, 112,
 113, 114, 125, 127, 134, 320, 322
 Luder, W. F., 475
 Ludewig, S., 598, 660
 Lund, E., 605, 660

M

Macallum, A. B., 18, 19, 26
 McAulay, J., 264, 321
 McDaniel, D. H., 475, 498, 547
 McDevit, W. F., 282, 322
 McGavin, S., 119, 122, 133
 MacInnes, D. A., 321, 392, 394, 398, 405,
 475, 558
 MacLennan, W. H., 188, 191
 McMeekin, T. L., 320, 322, 522, 548
 Malcolm, B. R., 119, 129, 133
 Margenau, H., 321
 Marsh, R. E., 105, 106, 109, 134
 Martell, A. E., 609, 638, 639, 660
 Martin, A. J. P., 66, 134
 Masket, A. V., 598, 660
 Mason, B., 4, 8, 9, 17, 25, 26

May, M., 453
 Maybury, R. H., 645, 661
 Mayer, J. E., 228, 239
 Mayer, M. G., 228, 239
 Mecham, D. K., 524, 548
 Meister, A., 211, 213, 240
 Meyerhof, O., 215, 240
 Michaelis, L., 645, 661
 Mihalyi, E., 135
 Miller, I. R., 542, 545, 548
 Miller, S. L., 12, 26
 Mills, E. A., 583, 585
 Mizushima, S., 643, 661
 Moffitt, W., 129, 134
 Moore, S., 53, 66, 67, 70, 73, 90, 92, 97,
 98, 134, 135
 Morawetz, H., 542, 547
 Morgan, J., 33, 34, 45
 Murphy, G. M., 321
 Murray, G. R., 643, 661

N

Näsänen, R., 453
 Naughton, M. A., 95, 134
 Neilands, J. B., 498, 548
 Neuberger, A., 60, 61, 89, 133, 135, 443,
 444, 445, 459, 460, 464, 467, 476, 494,
 495, 530, 548
 Neurath, H., 75, 132, 134
 North, A. C. T., 118, 119, 133
 Nozaki, Y., 464, 568, 627, 631, 632, 634,
 661

O

O'Brien, H., 571, 573, 575, 577, 590
 O'Brien, J. R. P., 540, 548
 Oneley, J. L., 372
 Onsager, L., 347ff., 385
 Oparin, A. I., 12, 16, 26
 Overbeek, J. T. G., 542, 548
 Owen, B. B., 291, 322, 405, 433, 441,
 448, 452(1, 12), 453, 472, 473, 475

P

Page, L., 321
 Palmer, A. H., 516, 517, 522, 523, 526,
 527, 529, 530, 534, 547
 Pals, D. T. F., 542, 548

Panofsky, W. K. H., 321
 Pardee, A. B., 240
 Parrish, R. G., 132, 134
 Partington, J. R., 385
 Parton, H. N., 452(3), 453
 Patterson, A., Jr., 558, 559, 560, 590
 Pauling, L., 26, 28, 29, 31, 32, 45, 84, 99,
 103, 104, 105, 106, 109, 110, 114, 115,
 134, 135, 232, 240, 407, 475, 553, 554,
 576, 590
 Pearson, L., 574, 585, 590
 Peck, E. R., 321
 Perlmann, G., 86, 135
 Perutz, M. F., 101, 132, 133, 134, 610
 Peters, J. P., 571, 589
 Phillips, M., 321
 Pinching, G. D., 452(2, 8, 10, 11), 453
 Pinsent, B. R. W., 574, 585, 590
 Planck, M., 239
 Portzehl, H., 542, 549
 Pound, A., 272, 273, 274, 321
 Powell, R., 205, 240
 Prentiss, S. S., 198, 240, 314, 322
 Putnam, F. W., 524, 548, 645, 661

Q

Quagliano, J. V., 643, 661

R

Raache, I. D., 95, 134
 Rabin, B. R., 643, 661
 Ramachandran, G. N., 118, 119, 135
 Randall, J. T., 118, 133
 Randall, M., 239, 274, 322
 Rands, D. G., 534, 536, 537, 549
 Rankama, K., 9, 21, 26, 550, 590
 Redfield, R. R., 83, 84, 86, 92, 93, 96, 97,
 98, 128, 129, 133, 135, 530, 547
 Ressler, C., 79, 133
 Rhinesmith, H. S., 84, 99, 135, 576, 590
 Rice, S. A., 542, 548
 Rich, A., 119, 120, 133, 135
 Richards, F. M., 320, 322
 Riggs, A. F., 64, 135
 Robbins, E. A., 215, 240
 Roberts, C. W., 79, 133
 Roberts, G. L., Jr., 464, 468, 476, 532,
 543, 549
 Robertson, J. M., 100, 135

Robinson, R. A., 46, 200, 240, 291, 292,
 322, 400, 404, 405, 433, 472, 473, 475
 Rollett, J. S., 126, 134
 Roos, P., 95, 134
 Rossini, F. D., 239
 Roughton, F. J. W., 560, 561, 571, 572,
 574, 577, 578, 580, 581, 582, 583, 584,
 585, 587, 589, 590
 Rubey, W. W., 15, 18, 26
 Rushbrooke, G. S., 228, 233, 239, 240
 Rykian, L. R., 465, 498, 548
 Ryle, A. P., 96, 99, 135

S

Sahama, T. G., 9, 21, 26, 550, 590
 Sanger, F., 81, 83, 84, 86, 89, 91, 95, 96,
 97, 133, 134, 135, 566, 590
 Scatchard, G., 198, 240, 264, 272, 281,
 295, 314, 321, 322, 519, 535, 548, 597,
 598, 600, 601, 603, 606, 613, 617, 619,
 646, 647, 649, 650, 651, 660, 661
 Scheinberg, I. H., 519, 535, 548, 597, 598,
 613, 646, 647, 649, 651, 661
 Schellman, J. A., 124, 125, 134, 135
 Scheraga, H. A., 510, 536, 545, 547, 548
 Schmidt, C. L. A., 56, 132, 465, 498, 548
 Schmitt, F. O., 118, 135
 Schnek, G., 464, 540, 548
 Scholes, S. R., 453, 561
 Schroeder, W. A., 84, 99, 135, 576, 590
 Schubert, J., 660
 Schubert, M. P., 483
 Schwarz, K., 271, 272, 281, 322
 Schwarzenbach, G., 458, 459, 462, 476,
 609, 638, 641, 642, 660, 661
 Scott, E. M., 205, 240
 Scott, R. L., 189, 240
 Sela, M., 542, 543, 544, 549
 Sen, D. N., 643, 661
 Sendroy, J., Jr., 558
 Shavit, N., 542, 548
 Shedlovsky, T., 558
 Shen, A. L., 603, 619, 650, 661
 Shepherd, R. G., 95, 134
 Shookhoff, M. W. S., 461, 474, 476
 Shore, H., 465
 Shore, W. S., 532, 533, 534, 535, 549
 Shugar, D., 464, 467, 476, 537, 549
 Sidgwick, N. V., 1, 24, 26
 Sillén, L. G., 638, 661

Simms, H. S., 488, 549
 Simpson, R. B., 645, 661
 Smith, E. R., 448
 Smith, E. R. B., 196, 197, 240, 309, 322,
 452(13, 14), 453
 Smith, L. F., 96, 135
 Smith, P. K., 196, 197, 240, 309, 322,
 452(13, 14), 453
 Smith, R. M., 213, 214, 239, 240, 467, 475
 Smyth, C. P., 341, 385
 Smyth, D. H., 205, 240
 Snell, E. E., 70, 135
 Sörensen, S. P. L., 605, 660
 Spackman, D. H., 73, 135
 Springall, H. D., 132
 Stadie, W. C., 571, 573, 575, 577, 590
 Stein, W. H., 53, 66, 67, 70, 73, 90, 92, 97,
 98, 134, 135
 Steinberg, D., 135
 Steinhardt, J., 538, 539, 549, 576, 590
 Stokes, R. H., 46, 200, 240, 291, 292, 322,
 400, 404, 405, 433, 472, 473, 475
 Straessle, R., 63, 135, 645, 661
 Street, A., 107, 133
 Stricks, W., 644, 645, 661
 Stuart, H. A., 385
 Sturtevant, J. M., 208, 240
 Sverdrup, H. V., 18, 26
 Swan, J. M., 79, 133
 Swanson, S. A., 532, 533, 534, 535, 536,
 549
 Szent-Györgyi, A. G., 130, 131, 133, 135

T

Tanaka, N., 645, 662
 Tanford, C., 464, 468, 476, 528, 529, 532,
 533, 534, 535, 536, 537, 539, 540, 543,
 546, 549
 Taylor, A. C., 452(13), 453
 Taylor, H. F. W., 467
 Taylor, J. F., 51, 135
 Thompson, A. R., 67, 90, 135
 Thompson, E. O. P., 67, 90, 135
 Thomson, D. W., 188, 191
 Tristram, G. R., 71, 73, 135, 527, 549
 Triwush, H., 652, 661
 Tupper, R., 588, 590
 Turba, F., 67, 90, 135

U

Urey, H. C., 26, 583, 585

V

Vallee, B. L., 587, 590, 620, 662
 van Goor, H., 587, 590
 Van Slyke, D. D., 571, 589
 van Vleck, J. H., 364, 385
 Velick, S. F., 598, 619, 620, 661
 Vickery, H. B., 56, 132
 von Muralt, A., 488, 548

W

Wagner, M. L., 534, 540, 549
 Wald, G., 54, 135
 Walker, F. M., 652, 661
 Warner, R. C., 320, 322, 530, 549
 Warren, B. E., 33, 34, 45
 Watts, R. W. E., 588, 590
 Waugh, D. F., 593, 660
 Weber, H. H., 542, 549
 Wells, A. F., 26, 46
 Westheimer, F. H., 461, 474, 476
 Wetlaufer, D. B., 274, 504
 Wilcox, P. E., 606, 634, 638, 639, 643, 644,
 660, 661
 Williams, R. J. P., 638, 662
 Wilson, J. N., 385
 Wissbrun, K. F., 558, 559, 560
 Wolbach, R. A., 64, 135
 Woods, H. J., 107, 108, 133
 Wormall, A., 588, 590
 Wyman, J., Jr., 371, 372, 377, 385, 525,
 539, 540, 549, 571, 578, 590, 635, 636,
 662
 Wynne-Jones, W. F. K., 452(9), 453

Y-Z

Yang, J. T., 129, 130, 134, 135, 536, 549
 Zaiser, E. M., 538, 539, 549, 576, 590
 Zucker, L., 504
 Zuffanti, S., 475
 Zweig, G., 67, 133
 Zwick, M., 542, 548

Subject Index

A

- ADP, *see* Adenosine diphosphate.
 ATP, *see* Adenosine triphosphate.
 Absolute temperature, *see under* Temperature.
 Acetaldehyde,
 dipole moment, 338
 gas, standard free energy of formation, 236
 Acetamide, dipole moment, 338
 Acetate, standard free energy of formation, 236
 Acetic acid,
 acid-base equilibria, 442-443
 acid dissociation constant, 407, 452
 liquid, standard free energy of formation, 236
 titration curve, 477
 Acetoacetate, standard free energy of formation, 236
 Acetoacetic acid, apparent dissociation constant, 466
 Acetone,
 dipole moment, 338
 salting out, 271
 solutions with chloroform, 188
 standard free energy of formation, 236
 Acetylene, salting-out constant, 274
 Acetylglycine, apparent dissociation constant, 466
 Acetylhistidine, dielectric properties, 372
 Acid(s) (*see also* Acid-base equilibria;
 Amino acids; & *under individual acids*),
 cationic, 416
 conductivity method for study, 421ff.
 conjugate acids & bases, 406-411, 415
 definitions, 406, 410-411
 electromotive force determinations, 429ff.
 hydrolysis of proteins, 56, 65
 indicator method, relative strength of acids, 423ff.
 nature, 406ff.
 nonaqueous media, 408
 polyvalent,
 equilibrium between different ionic forms, 510ff.
 general relations, 487-494
 relative strengths, determining, 421ff.
 spectrophotometric studies on acids & bases, 424-429, 499-502, 530-531, 537, 543-544
 strength, effect of varying dielectric constants, 471ff.
 effect of varying ionic strength, 441ff.
 structure, relation to acid strength, 454ff.
 table of acids & conjugate bases, 407
 tables of dissociation constants, 452-453, 458-459, 464-467, 495-496, 502-503
 Acid-base equilibria (*see also* Acids), 406-549
 electrostatic effects, 457ff., 512ff.
 hemoglobin, 537
 β -lactoglobulin, 522-530
 ovalbumin, 530-532
 ribonuclease, 536
 serum albumins, 532-536
 synthetic polyelectrolytes, including polypeptides, 540ff.
 Acid dissociation constants (*see also* Acids), 420ff.
 determination by conductivity, 421-423
 Ostwald method, 421-423
 thermodynamic functions, 450ff.
 effects of interionic forces, 422, 441-446
 of neighboring charged groups & dipoles, 457ff.
 tables of values & thermodynamic functions, 452-453, 464-467
 Acidity constants, *see* Acid dissociation constants.
 Aconitate, *cis*-aconitate, standard free energy of formation, 236

- Actin, lifetime, 49
- Activity, Activities,
 concept, 181
 distribution coefficients & solubilities,
 relations, 198ff.
 electrolytes, strong, 200
 ideal solution components, 186
 nonvolatile solute, determined from
 vapor pressure of solvent, 195
 standard states, 189-194
 thermodynamic, 181ff.
 vapor pressure, 191-194
- Activity coefficients,
 acids & bases, effects of ionic strength,
 441-446
 amino acids, 197
 decreased, with increased solubility,
 282
 dipolar ions, 296-320
 distribution coefficients & solubilities,
 relationships, 198ff.
 electrolyte solutions, 282ff.
 ionic interactions, effect, 282-296
 ionic strength, relationships, 282-320
 ions, effect of interionic forces, 282-296
 related to emf measurements, 432
 peptides, 197
 polyvalent acids (charged sphere
 model), 514ff.
 relation to different concentration
 units, 193-194
 salting-out effects, 266ff.
 standard states, 189-194
 vapor pressures, 191-194
- Adenine, apparent dissociation constant,
 467
- Adenosine, apparent dissociation constant,
 467
- Adenosine diphosphate,
 apparent dissociation constant, 467
 binding of cations, 216
 complexes, formation with cations,
 214-215
 formation from adenosine triphosphate,
 free energy, 210-216
- Adenosine-5'-phosphate, apparent dissociation
 constant, 467
- Adenosine triphosphate (ATP),
 apparent dissociation constant, 467
 binding of cations, 216
 formation of complexes, with cations,
 214-215
 free energy of hydrolysis, 210-217
- Adiabatic processes, 146ff.
 irreversible, 155ff.
 reversible, 148ff.
- Alanine,
 formation of carbamate, 574
- α -Alanine,
 acid dissociation constants, 452
 activity coefficient, 197
 constituent in important proteins, 72
 dielectric properties, 372
 partial molal volume, 173
 structure, 57ff.
- DL-alanine,
 copolymer with tyrosine, ionic
 strength, 544ff.
 solubility in water & ethanol, 306
 standard free energy of formation,
 236
- L-alanine,
 standard free energy of formation,
 236
- β -Alanine,
 acid dissociation constants, 453
 activity coefficient, 197
 dielectric properties, 372
 occurrence in nature, 57
 partial molal volume, 173
 solubility in water & ethanol, 306
- Alanyl peptides, apparent dissociation
 constants, 466
- DL-Alanylglycine, standard free energy
 of formation, 236
- Albumin, *see* Ovalbumin; Serum albumin.
- Alcohol dehydrogenase,
 binding of pyridine nucleotides, 619-
 620
 yeast, 619-620
 zinc as constituent, 6, 611, 620
- Aldolase, rabbit, N-terminal residues, 84
- Aliphatic acids, *see* Fatty acid ions; Fatty
 acids.
- Alkali, hydrolysis of proteins, 56
- Alkylammonium ions (*see also* Ionization
 constants),
 acid dissociation constants, 454

- Aluminum,
 biological significance, 5
 distribution in nature, 4, 11
 percentage in sea water, 21
- Amide N, constituent in important proteins, 73
- Amines, reaction with CO₂ to form carbamates, 571-578
- Amino acids,
 absolute spatial configuration, 60-61
 acid dissociation constants, 452, 464-467, 494ff.
 activity coefficients, 197
 analysis of proteins, 71-73, 81
 carbamate formation in, 572ff.
 collagen, 116-117
 C-terminal, 76, 83
 D-series, 60-61
 dielectric increments & electric moments, 367-377
 2,4-dinitrophenyl derivatives, 67, 82ff.
 equilibria of dipolar ions & uncharged isomers, 485ff.
 formation in primitive atmosphere, 13ff.
 fossils, discovery of amino acids as components, 16ff.
 history of chemical separation, 56
 hydroxyl-containing R groups, aliphatic & aromatic, 58
 ionization of side chains, 61ff.
 L-configuration of natural amino acids, 60-61
 lysine residues, 83
 melting points, 55ff.
 microbiological assay, 70
 N-terminal, 76, 82-84
 nonpolar aliphatic R groups, 58
 notation for residues, 76
 & peptides as dipolar ions, 54ff., 296-320, 367-377, 451ff., 485-487
 phosphate cross-links, 87
 protein hydrolysis products, 56
 reaction with CO₂ to form carbamates, 571-578
 separation, 57
 sequences in peptide chains, 93-96
 solubility, 306ff.
 spatial configuration, 57
 table of salting-out constants, 274
 of structure of side chains in proteins, 58-59
 uniqueness of α -amino acids in proteins, 57
- Aminobutyric acid,
 salting-out constant, 274
- α -Amino-*n*-butyric acid,
 acid dissociation constants, 452
 dielectric properties, 372
 DL- α -amino-*n*-butyric acid,
 solubility, 306
- β -Aminobutyric acid,
 dielectric properties, 372
- γ -Aminobutyric acid,
 acid dissociation constants, 453
 dielectric properties, 372
- DL- α -Amino-*n*-caproic acid,
 solubility, 306
- ϵ -Aminocaproic acid,
 acid dissociation constants, 453
 activity coefficient, 197
 dielectric properties, 372
 dipole moment, 297
 solubility, 306, 315
- π -Aminoheptadecylic acid betaine,
 dielectric properties, 372
 moment from dielectric increment,
 comparison, 377
- ζ -Aminoheptylic acid,
 dielectric properties, 372
- α -Aminoisobutyric acid,
 acid dissociation constants, 452
- ξ -Aminopentadecylic acid betaine,
 dielectric properties, 372
 moment from dielectric increment,
 comparison, 377
- Aminopeptidase, specificity, 74
- α -Aminovaleric acid,
 dielectric properties, 372
- γ -Aminovaleric acid,
 dielectric properties, 372
- δ -Aminovaleric acid,
 dielectric properties, 372
- Ammonia,
 absorption coefficient in water, 556
 acid-base equilibria, 407, 452
 comparison with other isoelectronic substances, 27ff.
 complexes with metal ions, 634, 640-641

- dipole moment, 338
 - gas, standard free energy of formation, 236
 - heat capacity, 38
 - hydrogen bond, compared with hydrogen bond in water, 36
 - metal complexes, 609
 - Ammonium group, influence of substituents on acid dissociation, 459
 - Ammonium ion,
 - acid dissociation constants, 407, 452
 - derivatives, acid dissociation constants, 452
 - standard free energy of formation, 236
 - Ampholytes, polyvalent (*see also* Acids; Polyvalent acids),
 - equilibrium between different ionic forms, 510ff.
 - Amphoteric substances, *see under* Dipolar ions.
 - Aniline,
 - acid-base equilibrium, 472
 - dipole moment, 338
 - Antigen-antibody reactions, 593
 - Arginine,
 - apparent dissociation constants, 465
 - constituent in important proteins, 72
 - guanidinium group, structure, 469
 - L-arginine, dielectric properties, 372
 - structure, 59
 - Argon,
 - biologically inert, 5
 - distribution in nature, 3-4
 - Arsenic, percentage in sea water, 21
 - Asparagine,
 - L-asparagine,
 - salting-in effect, 319
 - solubility, 199, 306, 319
 - salt solutions, 199
 - standard free energy of formation, 236
 - structure, 58
 - Aspartate,
 - L-aspartate, standard free energy of formation, 236
 - Aspartic acid,
 - acid dissociation constants, 453
 - constituent in important proteins, 73
 - copolymer with tyrosine, ionic strength, 544ff.
 - L-aspartic acid, standard free energy of formation, 236
 - structure, 58
 - Association constants,
 - dissociation constants, relationship, 519-520, 610-611
 - electrostatic effects, 645ff.
 - general discussion, 610-660
 - intrinsic microscopic constant, 624
 - metal ions & organic bases, relation to ionic structure, 642
 - polyvalent acids, 518ff.
 - protons, 518ff.
 - Association reactions,
 - competition effects, 651
 - electrostatic effects, 645ff.
 - general equation for binding a molecule or ion with n sites, 623-627
 - interaction effects, 626ff.
 - linked functions, 653ff.
 - reciprocal relations, 654ff.
 - statistical factors, 624
 - Atmosphere,
 - evolution, 12ff.
 - gases, distribution, 223
 - photosynthesis, 15
 - primitive, 12ff.
- B**
- Barium,
 - association with organic bases, 639
 - biological significance, 6
 - complexes with organic bases, 639
 - percentage in sea water, 21
 - Base(s),
 - binding to metallic ions, 626
 - conjugate, 410-411
 - definitions, 410-411
 - nature, 406ff.
 - table of acids & conjugate bases, 407
 - various meanings, 409
 - Basic dissociation constant, relation to acid dissociation, 410
 - Basicity constant, Brönsted's, definition, 518ff., 611
 - Beer's law, light absorption, & measurement of acid dissociation constants, 424ff.

- Benzbetaines (*o*, *m*, & *p*), dielectric properties, 372
- Benzene, dipole moment, 338
- Benzoic acid, acid dissociation constant, 452
- Benzoyl-L-tyrosine, 207-210
glycinamide, reaction with, enzymatic synthesis, 207
- Beryllium,
biological significance, 5
distribution in nature, 4ff.
- Betaine, activity coefficient, 197
- Bicarbonate ion,
acid-base equilibria, 453, 558ff.
standard free energy of formation, 236
- Bicarbonate solutions, carbon dioxide dissociation curves, 564
- Binding constants, *see* Association constants.
- Binding of ions to macromolecules, 521ff., 591-660
- Biochemical systems, *see* System.
- Biosynthesis, proteins, 49
- Bjerrum's theory, effect of substituents on strength of acids, 459ff.
- Blood,
carbamate functions in CO₂ transport, 577-578
carbon dioxide,
dissociation curves, 566
transported by blood, 561ff., 589
carbonic anhydrase, 562, 587ff.
hemoglobin as buffer, 589
plasma (*see also* Serum albumin), 27
binding of chloride ion, 650
ionic composition, 18-19
proteins, carbamate formation, 575
rates of reactions involving CO₂, 585-589
time required to flow through capillaries, 586
water content, 27
- Boltzmann constant, 226
- Boltzmann distribution law, 227, 286, 328ff., 332ff.
application to Debye-Hückel theory, 286
definition, 227
dielectric properties, 328ff.
dipole moment, relationship, 332ff., 353ff., 358ff.
mean total moment, calculation, 351-352
- Bonds, *see* Disulfide bonds; Hydrogen bonds.
- Boron,
biological significance, 5
distribution in nature, 4, 7
occurrence in sea water, 20
percentage in sea water, 21
- Bovine, *see under subheads* beef and cattle.
- Brand, Erwin, notation for amino acid residues, 76
- Bromine,
biological significance, 3
occurrence in sea water, 20
percentage in sea water, 21
- Bromoform, dipole moment, 338
- Brönsted, definition of acids & bases, 406
- Brönsted-Lowry theory of acids & bases, 409
- Buffer action, 417ff.
dibasic acids, 484
- Buffer values, 419-420
definition, 420
- Burk & Lineweaver, enzyme-substrate interactions, 622
- n*-Butanol, liquid, standard free energy of formation, 236
- Butyric acid & butyrate ion, standard free energy of formation, 236
- C
- Calcium,
biological significance, 5ff.
chelate complexes, 609
complexes with organic bases, 639
distribution in nature, 3-4, 9-10
percentage in sea water, 21
- Calcium ion, formation of complexes with casein, 598
binding to myosin, 599
- Calorie, definition, 145
- Caratheodory's formulation of Second Law of Thermodynamics, 140, 147-156

- Carbamate(s),
 formation from alanine, 574
 cysteic acid, 574
 from CO_2 & amines, 552ff., 571ff.
 with amino acids, "noncarbonate"
 equilibrium, 573ff.
 equilibrium constants, 574
 stable equilibrium, 574
 from glycine, 574
 with proteins, 575
 instability in acid form, 572ff.
 plasma proteins, carbamates in, 576ff.
- Carbamic acid (*see also* Carbamates),
 half-life, 572
 pK value, 572
 solubility of calcium & barium salts,
 572
- Carbamino compounds, *see* Carbamates.
- Cadmium ions, association constants
 with bases, 634
 complexes with organic bases, 639, 641
 interaction with serum albumin, 643
- Carbon,
 abundance, 8-10
 biological significance, 1-3, 5, 13-16,
 18, 21, 23-25
 distribution in nature, 4, 8ff.
 percentage in sea water, 21
- Carbon dioxide,
 absorption coefficients in water, 555ff.
 acid dissociation constants, 453, 557-
 561
 biochemical formation, 551
 biological significance, 2, 15ff., 25,
 550ff.
 blood dissociation curves, 566
 carbamate, formed from amines & CO_2 ,
 552ff.
 carbon dioxide, calculation of total
 content, in blood, 563-571
 carbonic acid, formed by hydration of
 CO_2 (*see also under main heading*,
 Carbonic acid), 550, 554, 557
 coenzyme A, in metabolism, 551
 dipole moment, 338
 dissociation curves, 561-571
 distribution in nature, 550ff.
 between water & air, 556-557
 efficiency as a buffer, 550ff., 558-571
 gas, standard free energy of formation,
 236
 hydration to carbonic acid, 550, 554,
 557, 583ff.
 rate, 578-589
 thermodynamics, 561
 molecule,
 linear shape, 30
 structure, 553
 photosynthesis, role, 551
 pyruvic acid, in metabolism, 551
 system, total carbon dioxide content,
 557
 uniqueness, 25
- Carbon monoxide, absorption coefficient
 in water, 556
- Carbonate ion, molecular structure, 553
- Carbonic acid,
 acid dissociation constants, 453-454
 dehydration, to carbon dioxide, 582-
 583
 heat of ionization, 560
 hydration from carbon dioxide, 551,
 554, 557-558
 rate, 579
 ionization, 235, 238, 453, 558ff.
 thermodynamic functions, for hydra-
 tion & ionization, 561
 Wien effect, 559
 ionization constant, 558
 second ionization constant, 453, 561
 molecule,
 structure, 553
 resonating structures, 554
- Carbonic anhydrase,
 biological function, 587
 blood, function in, 562
 contaminant of hemoglobin solutions,
 575
 isoelectric point, 588
 molecular weight, 588
 poisons for inhibition of, 588
 presence in pancreas, 588
 purification, 587-588
 sulfanilamide as inhibitor, 588
 zinc, constituent, 6, 588, 611
- Carbonyl-hemoglobin, *see* Hemoglobin,
 carboxyhemoglobin.
- Carboxyhemoglobin, *see under* Hemo-
 globin.

- Carboxyl groups (*see also* Acids),
 ionization, influence of substituents,
 table, 458
 tables of pK values, 452, 464-467, 495-496, 503
- Carboxylate ions, interactions with metal ions, 642-643
- Carboxylic acids, *see* Carboxyl groups; Acids.
- Carboxypeptidase, 83
 N-terminal residues, 84
 specificity, 74
 zinc as essential constituent, 6, 611
- Carotenoids, 53
- Cartesian coordinates, relation to polar coordinates, 245
- Casein,
 α -casein,
 phosphate cross-links, 87
 β -casein,
 phosphate cross-links, 87
 formation of complexes with calcium ion, 598
- Cationic acids, equation for pH , 416
- Cellulose, 48
- Centrifugal fields, systems in, 219-224
- Cesium, percentage in sea water, 21
- Charge(s),
 density, 293ff., 330
 electrostatic unit, 242
 image, *see* Image charges.
 proton unit, 28, 243
- Charged sphere,
 calculations, 265, 546ff.
 free energy, 252ff., 513ff.
 model,
 binding of anions by serum albumin, 647ff.
 electrical free energy, 513ff.
 limitations, 521-522
 modified model, with charges in specified locations, 546
 polybasic acids, 512ff.
 salting-out effect, relationship, 264
 self-potential, 253
- Chelate complexes, 608-609, 641-642
- Chemical elements,
 abundance, relative, 7ff.
 biological significance, 3-7
 distribution in nature, 3-12
 nuclear charges, 28ff.
 percentage in sea water, 21
- Chemical potentials (*see also* Activity; Activity coefficients),
 & activity, relationship, 182, 193-194
 interrelationships, 268ff.
 salting-out effect, relationship, 266ff.
- Chemical structure, biological significance, proteins, 81
- Chloride ions,
 association with mercaptalbumin, 651
 binding to serum albumin, 597ff., 646ff.
 in blood plasma, 18, 650
- Chlorine,
 biological significance, 5ff.
 distribution in nature, 3-4
 occurrence in sea water, 20
 percentage in sea water, 21
- Chloroacetic acid, acid dissociation constant, 452
- Chloroform,
 dipole moment, 338
 solutions with acetone, 188
- Chromatography,
 ion exchange resins, 67ff.
 paper, in amino acid analysis, 65
 proteins, 52-53
 ribonuclease, 53
- Chromium,
 complexes with bases, 626
 distribution in nature, 4
 nonessential in biochemical systems, 3, 5
- Chromosomes, nucleic acids, relationship, 48
- Chymotrypsin, α & β ,
 C-terminal & N-terminal residues, 84
 specificity, 74, 93
- Chymotrypsinogen, 84
 purification, 52
- Citrate ion,
 complexes with calcium, 609
 with metal ions, 639
 standard free energy of formation, 236
- Clark, W. M., use of terms "acid" & "base," 411
- Clausius-Clapeyron relation, thermodynamics, 164ff.
 equation, 165

- Clausius-Mossotti relation, 331ff.
& Debye theory, 331, 334
- Cobalt,
biological significance, 6
chelate complexes, 609
complexes with bases, 626, 639, 641
distribution in nature, 3-4ff.
percentage in sea water, 21
- Cobaltous complexes, 643ff.
- Coenzyme A in carbon dioxide metabolism, 551
- Cohn's equation for salting out, 275-276
- Cohn, *see* Cohn's equation for salting out
- amino acid analysis, 72-73
collagen I & collagen II, 122
denaturation, 123
electron microscopy, 118
heat shrinkage, 117
helical structure, 118ff., 122
metabolism, 49
molecular structure, 115ff.
diagram, 121
proline & hydroxyproline, 117, 123
molecular weight, 123
solubility, 118
solution, 123
tendon, 48
X-ray diffraction, 116ff.
- Competition effects,
association reactions, 651ff., 656
enzyme systems, 607-608
interactions of hemoglobin, 607-608
- Complexes (*see also* Association reactions),
protein with lipides, 52
- Components, system, *see* System, components.
- Conalbumin,
egg white,
N-terminal residues, 84
- Condenser, electrical, 379ff.
capacity, definition, 242
parallel plate, 329
- Conductance, *see* Conductivity.
- Conductivity,
electrolytes, 386-404
equivalent, 389-390, 401, 403-404
measurement of conductivity, 404
measurement of dielectric constants.
relationship, 384-385
- Ohm's law, 386
relaxation time, 401-402
resistance, 386-387
specific, definition, 387
table of ionic equivalent conductances, 400
- Conductors, 323ff.
dielectric constants, 323ff.
perfect, 324
- Conjugate acid, *see* Acid.
- Conjugate base, *see* Base.
- Coordinates, *see* Polar coordinates.
- Coordination numbers of metal ions, 626, 638
- Copolymers of tyrosine, 544
- Copper,
association constants with bases, 634, 639
binding to proteins, 592
biological significance, 4
chelate complexes, 609
complexes with bases, 626, 639
distribution in nature, 3, 6
percentage in sea water, 21
square configuration of coordinate bonds, 633
- Corticotropins, structure,
 α -corticotropin, sheep, 95
 β -corticotropin, hog, 94
structural relation to melanotropins, 95
- Coulomb's law of force between charged bodies, 241-242
- Creatine,
apparent dissociation constant, 465
dielectric properties, 372
standard free energy of formation, 236
- Creatinine, standard free energy of formation, 236
- Crystals,
close packing, 36
dipolar ions, 55
perfect, entropy, 228
proteins,
composition, 320
structure, 100, 131ff.
- Curie point (Pierre Curie), dielectric constants, relationship, 341ff.
- Cyanide ion, complexes with metal ions, 640

Cyclohexane, dipole moment, 338
Cysteic acid, formation of carbamate, 574
Cysteic acid groups, oxidation products of cystine, 85
Cysteine (*see also* Sulfhydryl groups),
 apparent dissociation constants, 465, 498–503
 constituent in important proteins, 63, 72
 interactions with metal ions, 644
 macroscopic ionization constants, 502ff.
 microscopic ionization constants, 496, 502–503
 notation of residues in protein structure, 76
 standard free energy of formation, 236
 structure, 59
Cysteine betaine (N-trimethylcysteine),
 apparent dissociation constant, 465
 pK value, 498
Cysteine ethyl ester, ionization constants, 502
Cysteinylcysteine, apparent dissociation constants, 465
Cysteinylglycine, ionization constants, 502
Cystine,
 constituent in important proteins, 72
 disulfide bonds, 48, 75ff., 85ff.
 disulfide linkages in β -keratin, 106ff.
 notation of residues in protein structure, 76
 salting-in effect, 316–319
 salting-out constant, 274
 salting-out effect, 316–318
 solubility, 316–318
 standard free energy of formation, 236
 structure, 59
Cystinyldidiglycine, moment from dielectric increment, comparison, 377
Cystinyldiglycine,
 dielectric properties, 372
 moment from dielectric increment, comparison, 377
Cytochromes, catalysts of oxidation, 53
Cytosine, apparent dissociation constant, 467

D

DNFB, *see* Dinitrofluorobenzene.
DNP, *see* Dinitrophenyl.
DPNH, *see* Diphosphopyridine nucleotide.
Debye-Hückel theory, ionic interactions, 249–251, 282–296, 401, 442ff.
 acid-base equilibria, 442ff., 512ff.
 activity coefficients, equations, 290ff.
 conductivity, 401
 Poisson's equation, relationship, 285ff.
Debye unit of electric moment, 337
Debye theory,
 Clausius-Mossotti relation, 331, 334
 dielectric constant, 328ff.
 applications, 335–341
 limits, 330
 modifications, 342
 salting-out effect, 274, 278–282
Degrees of freedom in phase equilibria, 176–179
Dehydration reactions, carbonic acid, 580
 velocity constants, for hydration & dehydration, 585
Denaturation, protein,
 collagen, 123
 hemoglobin, 537ff.
 & phenolic hydroxyl groups, 88–89, 531, 537
 proteins, 51, 64
 & sulfhydryl groups, 64
Density, electric charge, *see under* Charge, density.
Deuterium, exchange in peptides & proteins, 127–129
Dialysis, equilibrium, in study of molecular complexes, 594
 ϵ - ϵ' -Diamino-di(α -thio-*n*-caproic acid),
 dielectric properties, 372
 moment from dielectric increment, comparison, 377
Dibasic acids, 477ff.
 buffer action, 484
 dissociation constants, 481ff.
 equations for symmetrical, 482
 with independent groups, 483–484
 ionization, 477ff.
Dichlorobenzenes (*o*, *m*, & *p*), dipole moments, 338–339

- Dielectric constant(s), 323-385
 as auxiliary variable, 260ff.
 definition, 43, 242
 dipole moments, 323ff.
 effective, in acid-base equation, 461
 image charge problems, 258-263
 insulators, 323ff.
 ion-dipole interactions, 296-320
 ionic interactions, 282-296
 measurements, 378-385
 molecular polarity, 45
 polar liquids, 341ff.
 polar media,
 equations, 363
 relation of dielectric constants &
 polarization, 365-366
 salting-out effect, 263-282
 table of values, 44
 variations, effect on acid strengths,
 471ff.
 water, 43ff.
- Dielectric increment,
 definition, 369
 relation to dipole moments, 369-378
 relation to length of chain between
 charged groups, 374
 tables of values, 371-372, 377
- Dielectric sphere model, polyvalent
 acids, 546-547
- Diethyl ether,
 dipole moment, 338
 salting out, 263
- Diethylmalonic acid, ionization con-
 stants, 462
- Differential titrations, 654ff.
- Diglycine, *see* Glycylglycine.
- Diglycylcystine, moment from dielectric
 increment, comparison, 377
- ϵ - ϵ' -Diguanido-di(α -thio-*n*-caproic acid),
 dielectric properties, 372
 moment from dielectric increment,
 comparison, 377
- Diiodotyrosine,
 apparent dissociation constants, 464
 structure, 58
- Diketopiperazine, structure, 102
- Dimethyl ether, dipole moment, 338
- Dimethylamine, dipole moment, 338
- Dimethylammonium ion, acid dissoci-
 ation constants, 452
- 2,4-Dimethylimidazole, apparent disso-
 ciation constant, 464
- Dimethylurea,
sym-dimethylurea,
 dipole moment, 338
- Dinitrobenzenes (*m* & *p*) dipole mo-
 ments, 338
- 2,4-Dinitrofluorobenzene (DNFB), label-
 ing technique, terminal groups of
 proteins & peptide chains, 82, 93
- Dinitrophenyl (DNP),
 amino acids, 67, 82ff., 90
 lysine residues, 83
- Diphosphopyridine nucleotide (DPNH),
 interaction with alcohol dehydrogen-
 ase, 619
- Dipolar ions (*Zwitterions*), 54ff., 257ff.
 acid-base equilibria, 443ff., 451ff., 485-
 487
 activity coefficients in water & other
 media, 308ff.
 amino acids & peptides, 54
 amphoteric substances, 504ff.
 dielectric behavior, 369ff.
 ellipsoidal model, 304-305
 equilibria with uncharged isomers,
 485-487
 high dielectric constants of solutions,
 55, 368ff.
 hydrogen bond in crystals, 55
 melting points, 55
 partial molal volume, 173
 polar liquids, 364ff.
 salt, effect on solubility, 310-320
 salting-out effect, 278, 304-305
 solubility, 55, 306-320
 ethanol, 306
 water & other media, 306-310
 spherical model, 304
 surrounded by ion atmosphere, 297
 vibrational frequencies, 55
- Dipole(s) (*see also* Dipolar ions),
 electrical potential & energy, 254ff.
 Langevin function, for orientation,
 333ff.
 orientation in electric field, 330ff.
 point, definition, 257ff.
 torque on molecule in electric field, 254
- Dipole moment(s) (*see also* Dielectric in-
 crement; Dipoles),

- Debye unit, 337
definition, 254
determination, 337
dielectric constants, relation, 323-378
 Debye theory, 328ff.
 modification of, 342ff.
 Kirkwood theory, 356
 Onsager theory, 347
dipolar ions, 310-320, 369-377
flexible molecules, 340
& free rotation, 376
glycine, calculated from salting-in effect, 313-314
hydroxydecanoic acid polymers, 340
ion-dipole interactions, relationship, 296-320
molecular interactions, relationship, 357
per unit volume, 327, 353
salting-in constants, 314ff.
 table, 316
salting-in effect, 310ff.
 table, for important molecules, 338
vector additivity, 339
Dipole substituents, 462
Direct plot of association data, 614-615
Dissociation constants (*see also* Acid dissociation constants),
 association constants, relationship, 519-520
 dibasic acids, 481ff.
 n-valent acid, individual groups, 488
Distribution coefficients & solubilities, relation to activities, 198ff.
 in study of salting-out effect, 270-272
Disulfide bonds (*see also* Cystine, and *individual proteins*),
 cystine, 48, 75ff., 85ff.
 proteins, 48, 78, 96-87, 125ff.

E

- EMF, *see* Electromotive force.
Eadie's equation, enzyme-substrate relations, 622
Earth,
 atmosphere, *see* Atmosphere.
 crust, 8ff.
 history, 11
 structure, 8ff.
Edestin, salted in by sodium chloride, 320
Edman's reagent, *see* Phenylisothiocyanate.
Egg albumin, *see* Ovalbumin.
Elastic bodies, thermodynamics, 217
Electric displacement, 326ff., 345ff.
 dielectric constant, relationship, 326ff.
 electric intensity, 326ff.
Electric field, internal field strength, 329
Electric intensity, 329ff.
 dielectric constant, relationship, 326ff.
 electric displacement, relationship, 326ff.
 flux & Gauss's law, 246ff.
Electric moment, *see* Dipole moment.
Electrical conduction, *see* Conductivity.
Electrical energy of ion, effect of surrounding ions, 289-290
Electrical free energy of charged sphere, 264ff., 513ff., 647ff.
Electrical mobility, *see* Mobility, electrical.
Electrical quantities, 241ff., 323ff.
 dimensions, 387
Electrical susceptibility, 327ff.
Electrolytes, 386-404
 activity, 200-201
 conductivity, 386-404
 solutions, activity coefficients, 290ff.
 strong, electrical conductivity, 391-392
 weak, electrical conductivity, 391-392
Electromotive force measurements,
 determination of binding of chloride & other ions, 596ff.
 pH determination, 429ff.
Electron microscopy, collagen, 118
Electronic polarizability, 337
Electrophoresis, study of molecular complexes, 599
 & ionic mobilities, 399-401
Electrophoretic effect, ion conductivity, 401-402
Electrostatic effects on association reactions, 645ff.
 on polybasic acids, 512
Electrostatic unit of charge, 242ff.
Electrostatics, 241-322
 dielectric constants, 323ff.
Elements, *see under* Chemical elements; and *individual elements*.

- Ellipsoid method,
 measuring dielectric constants, 379
- Energy (*see also* Free energy; Work),
 dipolar ion, calculation in ionic solution, 300-303
 dipole, 297
 equations relating to other quantities, 164-168
 First Law of Thermodynamics, 142-146
 levels in molecules, & thermodynamic functions, 227
- Enthalpy, *see* Heat content.
- Entropy, 140, 148ff., 155ff., 162ff., 224ff., 455ff.,
 calculation from heat capacity data, 233-234
 factors in association reactions, 642
 heat capacity, 233-234, 456
 of ionization,
 & acid-base equilibria, 455ff., 463
 language entropy, information theory, 226
 mixing, ideal solutions, 186, 229ff.
 perfect crystal, application of Third Law of Thermodynamics, 228
 phenolic hydroxyl groups, ionization, 463, 533
 probability, time, & entropy, 159
 randomness, 225ff.
 residual, of ice, 231
 standard, ionization, 451ff.
 statistical interpretation, 224-228
 vaporization of water, 167ff.
- Enzyme kinetics, relation of equations to those for binding, 620
- Enzymes (*see also individual enzymes*),
 activity, 128-129
 hydrolysis of proteins, 56, 92
 proteolytic, specificity, 74-75
- Epidermin, structural relation to keratin, 115
- Equilibrium (*see also* Acid dissociation constants),
 between different ionic forms, polyvalent acids & ampholytes, 510
 constants, 201ff., 557ff.
 heat of reaction, 206
 for hydration of CO_2 , 557ff., 585
 pH, effect, in malate-fumarate equilibrium, 206
 pressure, effect, 203
 criteria, 160-161
 dialysis, & molecular complex formation, 594
 free energy & maximum work, 159ff.
 malate-fumarate, 204-207, 235
 peptide bond synthesis, 207-210, 238
 temperature effect, 203-204
- Equipotential surfaces,
 definition, 244
 image charges, 259ff.
- Esters, effect on ionization of neighboring acid group, 485
- Ethane, dipole moment, 338
- Ethanol, liquid, standard free energy of formation, 236
- Ethanolammonium ion, acid dissociation constants, 452
- Ethyl acetate,
 dipole moment, 338
 salting out, 272, 274
- Ethylenediamine,
 chelate complexes with metals, 609, 639-641
 table of acids & conjugate bases, 407
- Ethylenediaminetetraacetate, chelate complexes with metals, 609
- External moment, dipole, 350
- F**
- Falkenhagen theory, conductivity, 401ff.
- Faraday equivalent of electric charge, 390, 431
- Faraday's law of electrolytic conduction, 388-389
- Fatty acid anions, binding to serum albumins, 89
- Fatty acids, acid dissociation constants, 454
- Ferric ions, interaction with sulfhydryl compounds, 645
 hydrated, as acids, 409
- Ferrous ions, interaction with sulfhydryl compounds, 645
- Fibrinogen,
 salting-out constant, 274
 structural relation to keratin, 115
- Fibroin, silk, structure, 104-106

Flexible molecules, dipole moments, 340

Fluorine,
 biological significance, 5
 distribution in nature, 4
 percentage in sea water, 21

Force between two point charges, 378

Formaldehyde,
 dipole moment, 338
 gas, standard free energy of formation, 236
 titration curves of amino acids & proteins, effect of added formaldehyde, 524

Formation function (J. Bjerrum), 615

Formic acid,
 acid dissociation constant, 452
 dipole moment, 338

Free energy (*see also* Energy; Chemical potential; Work),
 activity, relationship, 181ff.
 charged spherical ion, 263ff., 512ff., 645ff.
 chemical potentials, 174-175
 definition, 159-162
 elastic bodies, 217-219
 equilibrium, relationship, 159-162, 201ff.
 gravitational & centrifugal fields, 219-224
 of ionization, effect of dielectric constant, 473
 tables of values, 452-453, 464-467
 mass law, 201-217
 standard, of formation, 232-234
 table of values, 236-237

Free rotation, dipole moments, 374ff.

Frictional coefficient, 391

Fructose, standard free energy of formation, 236

Fumarate,
 equilibrium with malate, 204-207, 235
 standard free energy of formation, 236

Fumaric acid, standard free energy of formation, 236

Fusion, heat of, *see* Heat of fusion.

G

α -D-Galactose, standard free energy of formation, 236

Gases,
 perfect, in atmosphere, 223ff.
 salting-out constants, 274

Gauss's law, 246ff., 323ff.
 derivation, 246ff.

Gibbs-Duhem relation, 174-175, 186ff., 195
 equation, 195
 extended form, 220

Gibbs free energy, *see* Energy, free.

Glass electrode, calibration of pH measurements, 449ff.

Globin, relation to heme, 52ff.

Glucagon, pancreas hormone, structural formula, 94, 96

Glucose,
 conversion to glucose-6-phosphate, 215
 formation from glucose-6-phosphate, 215

α -D-Glucose, standard free energy of formation, 236

Glucose-6-phosphate, equilibrium constant in hydrolysis, 215

Glucose 1-phosphoric acid, acid dissociation constants, 452

Glutamate, conversion to glutamine, 210-214

L-Glutamate, standard free energy of formation, 236

Glutamic acid,
 constituent in important proteins, 73
 dielectric properties, 372
 ionization constants, 452, 495-496

L-Glutamic acid, standard free energy of formation,
 microscopic & macroscopic constants, pK values, 494ff.
 structure, 58

Glutaminase, catalyst in glutamine synthesis, 212-214

Glutamine,
 conversion from glutamate, 210-214
 structure, 58
 synthesis from glutamate & ammonium ion, 211ff.

L- γ -Glutamylcysteine, ionization constants, 502

Glutathione,
 interactions with metal ions, 644
 ionization constants, 465, 502

- Glycerol, liquid, standard free energy of formation, 236
- Glycerol 2-phosphoric acid, acid dissociation constants, 452
- Glycinamide, 207-210
 benzoyl-L-tyrosine reaction, enzymatic synthesis, 207
- Glycine (*see also* Polyglycine),
 acid dissociation constants 452, 485-487
 activity coefficient, 197
 ethanol-water solutions, 309ff.
 salt solutions, 310ff.
 water solutions, 309ff.
 combination with metal ions, 638
 constituent in important proteins, 72
 dielectric behavior, 368-372
 dielectric increment, molar, of dipolar ion, 369-372
 dipole moment, 296ff., 369-372
 calculated from salting-in effect, 313-314
 equilibria of dipolar ions & uncharged isomers, 485ff.
 formation of carbamate, 574
 ionization constants in various media, 473
 isoelectric point, 506-507
 metal complexes, 609
 partial molal volume, 173, 369-372
 pH of solution, 507
 polarizability, 371
 salting-in effect, 313-314, 319
 salting-out effect, 313ff.
 solubility, 278, 306, 310ff., 319
 ethanol-water solutions, 306ff.
 salt solutions, 310ff.
 water solutions, 306ff.
 space model, 296ff.
 standard free energy of formation, 236-237
 structure, 58
 table of acids & conjugate bases, 407
 titration curve, 477ff.
- Glycine betaine, dielectric properties, 372
- Glycine dipeptide, *see* Glycylglycine.
- Glycine heptapeptide, dielectric properties, 372
- Glycine hexapeptide, dielectric properties, 372
- Glycine pentapeptide, dielectric properties, 372
- Glycine tetrapeptide, dielectric properties, 372
- Glycine tripeptide, dielectric properties, 372
- Glycogen, 48
 standard free energy of formation, per glucose unit, 236
- Glycolamide,
 partial molal volume, 173
 solubility,
 water & ethanol, 306
- Glycolate ion, standard free energy of formation, 236
- Glycolic acid, acid dissociation constant, 452, 498
- Glycyl peptides,
 apparent dissociation constants, 466
 dielectric properties, 372
 solubilities, 306ff.
- Glycylalanine, activity coefficient, 197
- Glycylglycine,
 acid dissociation constants, 453
 activity coefficient, 197
 dielectric properties, 372
 interaction with metal ions, 643-644
 salting in, 315
 solubility, water and ethanol, 306
- Glycylphenylalanine, dielectric properties, 372
- Glycyl-L-tyrosine, ionization constants, 426-429
- Glyoxylate, standard free energy of formation, 237
- Gramicidin S, antibiotic, 78
- Gravitational fields, thermodynamic effects, 219-224
- Growth hormone, *see* Somatotropin.
- Guanidine,
 apparent dissociation constant, 465
 derivatives, ionization constants, 469
- Guanine, apparent dissociation constant, 467
- Guanosine, apparent dissociation constant, 467
- Güntelberg-Müller charging process, 290

H

- Heat (*see also* Heat of ionization),
 chemical reactions, relation to equilibrium constant, 203
 definition, 144
 relation to entropy, 148ff.
 & volume changes in water-propanol mixtures, 192
- Heat capacities,
 changes associated with acid dissociation constants, 457
 changes on ionization, 456
 entropy calculation, 233-234, 456
 partial molal, 173
 temperature effect, in chemical reactions, 204
 thermodynamic relations, 165
 water & other substances, 37
- Heat content, enthalpy, definition, 162
- Heat of fusion,
 latent, 39ff.
 table of heats of fusion, 40
- Heat of ionization, 455
 acidic groups in proteins, relation to nature of group, 525, 532-533
 imidazole groups, 525-526
- Heat, specific, *see* Specific heat.
- Heat of vaporization,
 hexane, comparison with water, 39
 hydrogen fluoride, comparison with water, 28
 isoelectronic sequences of hydride molecules, 28
 table of values for water and other substances, 39
 water, comparison with other substances, 28ff., 39ff.
- Helical cables, ropes & other structures, *see under* Peptides, chains.
- Helical & nonhelical regions in globular proteins (*see also* Peptides), 129
- Helical structures,
 collagen, 122
 distinction between right- & left-handed forms, 111, 129-131
 insulin, 127-128
 proline residues as influence on stability, 125
 stability, 123-125
- Helium,
 abundance, 8-10
 biologically inert, 5
 distribution in nature, 4, 8
- α -Helix (*see also* Helical structures), 110ff.
- π -Helix, 112ff.
- Heme-linked acid groups, *see under* Hemoglobin.
- Heme proteins, 610
- Hemocyanins, 610
 association & dissociation reactions, 593
- Hemoglobin,
 acid denaturation, 576
 amino acid analysis, 71ff.
 binding of water molecules to iron atoms, 605
 carbamate formation, 575-576
 carbon dioxide transport, function, 562, 575
 carboxyhemoglobin, 318, 400-401
 horse,
 solubility, 318
 human,
 ion mobility, 400-401
 combination with oxygen & carbon monoxide, 607-608, 635
 denaturation, in acid solutions, 537
 equilibrium between different ionic forms, 511
 heme-linked acid groups, 540, 569ff., 608, 653
 horse, 274, 318, 372
 dielectric properties, 372
 salting-out constant, 274
 solubility, 318
 human, 274, 400-401
 salting-out constant, 274
 interaction with oxygen, 591ff.
 interactions of heme groups, 606-608, 635
 iron porphyrin group, attachment, 52
 lifetime in red cells, 49
 molecular shape, 132
 molecular structure, 99
 N-terminal residues of various species, 84
 peptide chains, 88
 relation to carbon dissociation curves, 566

- salting-in effect, 319
- separating out carbonic anhydrase, 587ff.
- solubility, 318
- sulfhydryl groups, 63, 88
- terminal amino groups, 565ff., 576
- titratable groups, release, in acid solutions, 538
- titration curve, 525-526, 537-540
- Henderson-Hasselbalch equation, 418
- Henry's law, 186ff.
- Histidine,
 - apparent dissociation constants, 464
 - constituent in important proteins, 73
 - interactions with metal ions, 644
 - structure, 59
- Hittorf method, determining transference numbers, 394-395
- Hofmeister series, relation to salting-out effect, 277
- Homocysteine, ionization constants, 502
- Hormones (*see also* Insulin),
 - corticotropins, 94-96
 - glucagon, 94-96
 - melanotropins, 94-96
 - oxytocin, 78
 - posterior pituitary, 78-79
 - vasopressin, 79
- Hydration of carbon dioxide,
 - carbonic anhydrase, effects on, 587-590
 - catalytic effects, 584
 - equilibrium constants, 558, 585
 - velocity constants, for hydration & dehydration, table, 585
- Hydride molecules (water, ammonia, methane, etc.),
 - boiling points, 29ff.
 - heats of vaporization, 29ff.
 - melting points, 29ff.
- Hydrofluoric acid, *see* Hydrogen fluoride.
- Hydrogen (*see also* Hydrogen bonds),
 - absorption coefficient in water, 556
 - abundance, 8-10
 - biological significance, 3-5, 11-18, 20, 23-25
 - distribution in nature, 4, 7-11
- Hydrogen bonds, 31ff.
 - collagen, 121-123
 - cross-links in protein structure, 88
 - dipolar ion crystals, 55
 - effect on titration curves of polyvalent acids, 545
 - energy, in water, 32ff.
 - helical structures, 110ff., 124
 - influence on ionization constants, 532, 545
 - peptide chains, 102-103, 106ff.
 - phenolic hydroxyl groups, 62, 88, 467, 531
 - polyglycine II, 120-121
 - strength, 124-125
 - tetrahedral arrangement in water & ice, 36
 - water structure, 31ff.
- Hydrogen chloride, dipole moment, 338
- Hydrogen cyanide (*see also* Cyanide ion), salting in, 271
- Hydrogen electrode, 430ff.
- Hydrogen fluoride,
 - comparison with other isoelectronic substances, 28ff.
 - hydrogen bond compared with hydrogen bond in water, 36
- Hydrogen ion (*see also* Acids),
 - electrical mobility, 35, 400
 - hydronium ion, oxonium ion, 413-414
 - oxonium ion, commonly referred to as hydrogen ion, 413
 - standard free energy of formation, 236
 - table of acids & conjugate bases, 407
- Hydrogen ion activity, *pH*
 - calculation, 414ff.
 - concept, 413-414, 437ff.
 - measurement, 437-450
- Hydrogen peroxide, dipole moment, 338
- Hydrogen sulfide,
 - gas, standard free energy of formation, 237
 - hydrogen bond in, compared with hydrogen bond in water, 36
 - salting-out constant, 274
- Hydrolysis, proteins with acids, alkali, & enzymes, 56
- Hydronium ion, *see* Hydrogen ion.
- Hydroxybenzoic acids (*m* & *p*), apparent dissociation constants, 464
- β -Hydroxybutyrate, standard free energy of formation, 237

- β -Hydroxybutyric acid, standard free energy of formation, 237
- Hydroxydecanoic acid, polymers, dipole moments, 340 polarizability, 340
- Hydroxyl ion, standard free energy of formation, 237
- Hydroxylysine, constituent in collagen, 73, 116-117 structure, 59, 61
- Hydroxyproline, acid dissociation constants, 452 activity coefficient, 197 constituent in collagen, 73, 116-117, 122-123 hydroxy-L-proline, 61 structure, 59
- Hypsometric equation, distribution of components, 221ff. law, 222ff.
- I
- Ice, entropy, residual, 231ff. expansion of water in ice formation, 40-41 heat of fusion, 33 phase diagram, 178 structure, 32 various forms, 41, 178 water molecules, tetrahedral coordination, 31ff.
- Image charges, 258-263, 346 in dielectric media, 260-264 salting-out effect, 281
- Imidazole, apparent dissociation constants, 464, 636-637
- Imidazole groups, carbon dioxide dissociation curves, 566ff. complexes with metal ions, 634 ionizable, proteins, *pK* values, tables, 527, 534, 536 ionization constants, 464, 468 in serum albumin, interaction with zinc ions, 642
- Imino acid R groups, structure, 59
- Immunological reactions, criterion of protein purity, 51 molecular interactions, 593
- Impedance, electrical, dielectric constant measurements, relationship, 380ff.
- Independent groups, binding of molecules & ions, 610ff.
- Indicator method, relative strengths of acids, 423ff.
- Inductance of electrical circuits, 379-383
- Inductive effect, ionization constants, 457-463
- Infrared spectra, dipolar ions, 55
- Insulating sphere, effect on field in dielectric media, 261-263
- Insulators, dielectric constants, 323ff.
- Insulin, A & B chains, 86 amino acid analysis, 71ff., 76 association reactions, 593 beef, amino acid composition, 72-73 oxidation by performic acid, 86 beef, hog, horse, sheep, whale: structural formula, 94 disulfide bonds, 86, 96, 125ff. helical structure, possible, 127-128, 130 molecular structure, 97 molecular weight, 86 phenolic hydroxyl groups, 62, 88ff. stereochemistry, 126-128 titration curve, 540
- Integrating factors, 150ff.
- Interactions, association reactions, 608 binding by an equivalent set of groups, 626 hemoglobin, 608 linked functions, 653ff. metal imidazole complexes, 631ff. positive & negative, 625-626, 636 reciprocal relations, 608 serum albumin, 608
- Interionic forces (*see also* Debye-Hückel theory), effect of conductivity, 401-404 relation to molecular complex formation, 594
- Internal field (*F*), calculation in dielectric media, 329ff., 346, 351ff.

- Intrinsic constants,
 individual group constants, 481ff.
 intrinsic microscopic association
 constants, 630
 n-valent acid's individual group, 488
Iodide ions, association with mercaptal-
 bumin, 651
Iodine,
 biological significance, 3
 occurrence in sea water, 20
 percentage in sea water, 21
Iodoacetamide, reaction with sulfhydryl
 groups, 85
Iodoform, dipole moment, 338
Ion (*see also* Debye-Hückel theory;
 Salting-out effects),
 binding of ions to macromolecules,
 521, 594ff., 645ff.
 to proteins, calculated from *pH*
 shifts, 601-603
 calculation of electrical energy due to
 surrounding ions, 289-290
 potential, 285ff.
Ion conductances,
 definition, 393
 table of values, 400
Ion-dipole interactions, 296-305
Ion exchange resins,
 chromatography, 67ff.
 electrodes, in study of protein inter-
 actions, 650-651
 role in establishing criteria of purity, 52
 in removing ions from protein solu-
 tions, 600
 separation of proteins, 52
Ion mobilities, 390-392, 399ff.
 ion velocities, 390
Ion potential, notation, 285
Ionic equivalent conductances, table, 400
Ionic interactions, 282ff.
Ionic solutions, 241, 263ff., 386ff.
Ionic strength,
 definition, 273
 effect on titration curves of polyvalent
 acids, 512ff.
 relation to activity coefficients,
 of ions, 282-296
 of dipolar, 296-320
 solubility of dipolar ions, effect, 310-
 320
Ionization, *see* Acids; Acid dissociation
 constants.
Iron (*see also* Hemoglobin),
 abundance, 8-10
 acid properties, ferric ion, 409
 binding to proteins, 592
 biological significance, 6ff.
 complexes with bases, 626
 distribution in nature, 3-4, 8
 percentage in sea water, 21
Irreversibility, 146ff.
 irreversible processes, 146ff., 157ff.
 thermodynamic concepts, 217
 relation to Second Law of Thermo-
 dynamics, 146-147
Isocitrate ion, standard free energy of
 formation, 237
Isoelectric points,
 amino acids, simple, 506-507
 amphoteric substances, 504ff.
 isoionic solutions, relationship, 603ff.
 polyvalent ampholytes, 506-510
Isoelectric solutions, relation to isoionic
 solutions, 599, 603ff.
Isoelectronic substances, related to water,
 physical properties, 27ff.
Isoionic & isoelectric solutions, 599, 603ff.
 mercaptalbumin, 651
Isoionic point (*see also* Isoelectric points),
 definition, 600
Isoleucine,
 constituent in important proteins, 72
 structure, 58, 61
Isopiestic vapor pressure method, activ-
 ity coefficients, 196
Isopropanol, standard free energy of
 formation, 237
- K
- Keratin, 48,
 α -keratin,
 α -helix structure & other possible
 keratin structures, 110ff.
 molecular structure, 110ff.
 β -keratin,
 cystine disulfide linkages, 106ff.
 molecular structure, 106ff.
 X-ray diffraction pattern, 107ff.
 cystine, cross-links, 86
 wool, amino acid analysis, 72-73

Keto group, effect of acidity of neighboring groups, 466
 α -Ketoglutarate,
 formation from oxalosuccinate, 552
 standard free energy of formation, 237
Kirkwood's theory, ion-dipole interactions, 299, 311ff.
 dielectrics, 347ff.
 polar liquids, dielectric constants, 356ff.
 water, calculation of dielectric constant, 367

Kirkwood & Westheimer's theory,
 effect of substituents on acid strength, 461ff.

Kohlrausch's law, ion conductance, 393

L

Lactamide,
 partial molal volume, 173
 solubility, water & ethanol, 306
Lactate, standard free energy of formation, 237

Lactic acid, acid dissociation constant, 452

β -Lactoglobulin, 526
 amino acid analysis, 71-73
 crystals, composition, 320
 dielectric properties, 372
 heat of ionization, 526
 ionic strength, effect on titration curve, 517, 523ff.
 ionizable groups, table, 527
 molecular size, 72, 528
 salting-in effect, 319
 salting-out constant, 274
 titration curve, 522ff.
 effect of ionic strength, 523

α -Lactose, standard free energy of formation, 237

β -Lactose, standard free energy of formation, 237

Langevin function, dipole orientation, 333ff., 353, 359

Language entropy, information theory, 226

Laplace's equation, 251ff., 325, 349-350
 derivation, 251ff.
 solutions, 344ff.

Lead ions (Pb^{II}),
 complexes, with organic bases, 639
 with serum albumin, 643

Legendre function,
 associated spherical harmonics, 344-345
 ordinary, 345

Leucine,
 constituent in important proteins, 72
 DL-leucine,
 standard free energy of formation, 237

 L-leucine,
 solubility, 306
 standard free energy of formation, 237

 salting-out constant, 274

DL-Leucylglycine, standard free energy of formation, 237

L-Leucylglycylglycine, dielectric properties, 372

Levulinic acid, apparent dissociation constant, 466

Lewis, G. N., theory of acids & bases, 410

Life, origin on earth, 15ff.

Ligands, *see* Association constants; Association reactions.

Limiting law, for equivalent conductance, 403-404

Linderstrøm-Lang equation for polyvalent acids, 510ff.

Linked functions in association reactions, 653ff.

Lipides, complexes with protein, 52

Liquid junction potential, 439ff.

Lithium,
 biological significance, 5
 distribution in nature, 4
 percentage in sea water, 21

Lithosphere of earth, 11

Lysine,
 apparent dissociation constants, 465
 constituent in important proteins, 73
 copolymers with tyrosine, titration curves, 544

β -lactoglobulin, 64
 residues, dinitrophenyl derivatives, 83
 structure, 59

- Lysozyme, egg white,
 amino acid analysis, 71-73
 C-terminal and N-terminal residues,
 84
 isoelectric point, highly alkaline, 71
 titration curve, 540
- Lysylglutamic acid,
 dielectric properties, 372
 moment from dielectric increments,
 comparison, 377
- Lysyl-lysine, apparent dissociation constants, 465
- M**
- MSH, *see* Melanotropin.
- Macallum, A. B., hypothesis of composition of sea water and plasma, 18-19
- Macromolecules,
 binding of protons, 519-521
 biological significance, 47ff.
- Macroscopic & microscopic *pK* values,
 aminothiols, certain, 502
 L-cysteine, 502
 L-cysteine ethyl ester, 502
 L-cysteinylglycine, 502
 glutamic acid & its esters, table, 495
 glutathione, 502
 β -mercaptoethylamine, 502
 DL-homocysteine, 502
 theoretical considerations, 487-494
- Macroscopic state, microscopic configurations as constituents, 225-226
- Magnesium,
 abundance, 8-10
 biological significance, 5
 combination with organic bases, 639
 complexes with organic bases, 639
 distribution in nature, 3-4, 8
 ions,
 binding to adenosine diphosphate & adenosine triphosphate, 216
 percentage in sea water, 21
- Malate,
 equilibrium with fumarate, 204-207, 235
 standard free energy of formation, 237
- Malonate ion, *pK* value, 407
- Malonic acid, ionization constants, 407, 461-462
- Maltose, β -maltose, standard free energy of formation, 237
- Manganese,
 binding to proteins, 592
 biological significance, 5
 complexes with organic bases, 639
 distribution in nature, 4
 percentage in sea water, 21
- Manganous complexes, 643ff.
- Mannitol, standard free energy of formation, 237
- Mass law, derivation and applications, 201-217
- Maxwell's equations, electrodynamics, 157, 323
- Melanocyte-stimulating hormone, *see* Melanotropin.
- Melanotropin(MSH),
 hog and beef, structures, 95
 structural relation to corticotropins, 95
- Melting points,
 amino acids and peptides, 55
 water & other hydrides, 28ff., 40
- β -Mercaptoethylamine, ionization constants, 465, 502
- Mercuric ion, complexes with bases, 638
- Mercury,
 association constants with organic bases, 634, 639
 interaction with serum mercaptalbumin, 63-64, 645
- Meromyosin, helical configurations, 130-131
- Metal ions,
 association with bases, tables of values, 634, 639, 641
 association with proteins, 521, 598-599, 642-643
 calculation of association constants, 626ff.
 chelate complexes, 608-610, 643-645
 coordination numbers, 626
- Metalloporphyrins as chelate complexes, 609-610
- Methane,
 comparison with other isoelectronic substances, 27-28ff.

- dipole moment, 338
- molecule, electronic structure, 30
- Methanol, liquid, standard free energy of formation, 237
- Methionine,
 - constituent in important proteins, 72
 - oxidation by performic acid, 85
 - structure, 59
- Methyl acetylene, dipole moment, 338
- Methyl alcohol, dipole moment, 338
- Methyl bromide, dipole moment, 338
- Methyl chloride, dipole moment, 338
- Methyl iodide, dipole moment, 338
- O-Methyl tyrosine, microscopic ionization constants, 504
- Methylamine, dipole moment, 338
- Methylammonium ion,
 - acid dissociation constants, 407, 452
 - titration curve, 477
- S-Methylcysteine, pK value, 498
- 4-Methylimidazole,
 - apparent dissociation constants, 464
 - carbon dioxide dissociation curves, 566ff.
 - complexes with copper & zinc, 626-634
- Methylthioglycolate, apparent dissociation constants, 465
- Michaelis constant, enzyme-substrate interaction, 621-623
- Microbiological assay, amino acid analysis, 70ff.
- Microscopic configurations, thermodynamic systems, 225-226
- Microscopic ionization constants,
 - cysteine, 500ff.
 - glutamic acid, 393ff.
 - polybasic acids, different forms, 489fi.
 - sulfhydryl groups, 500ff.
 - tyrosine, 504
- Mobilities of ions, *see* Ion mobilities.
- Mobility, electrical,
 - ion conductances, relationship, 390ff., 399
 - isoelectric points of ampholytes, relationship, 505
- Molal quantities, partial, 170-174
- Molal volumes, partial, 171ff.
- Molar dielectric increment, *see* Dielectric increment.
- Molecular complexes (*see also* Association constants),
 - general consideration of formation, 605-606
 - interaction of bound groups, 605-606, 626ff., 653ff.
 - law of mass action, 593
 - methods of measuring, 594ff.
 - proteins, 50
 - spectroscopic studies, 596
 - study by electrophoresis, 599
 - by equilibrium dialysis, 594
 - by observing pH changes when salts are added, 599ff.
 - by ultracentrifuge, 598, 619
- Molecular interactions, general aspects, 591-660
- Molecular properties, *see under* Molecules.
- Molecular rearrangements, dissociation of acids, relationship, 482-483
- Molecular weights, 80, 223
 - determination in ultracentrifuge, 223
 - insulin, 86
 - proteins, 72, 80
- Molecules (*see also individual molecules*),
 - energy levels, 227
 - flexible, dipole moments, 340
 - nonpolar properties, 257ff.
 - orientation, 332ff.
 - polar, 241ff.
 - polarization, dielectric, 328ff.
 - properties,
 - dielectric constants, relationship, 324ff.
 - polarization, dielectric, relationship, 328ff.
 - structure, carbon dioxide, 553
 - water, 27-45
 - thermodynamic functions, 227
- Molybdenum,
 - biological significance, 3
 - percentage in sea water, 21
- Moving boundary method of determining transference numbers, 395-401
- Muscular contraction, possible mechanism, 542
- Myoglobin,
 - association with oxygen, 615
 - horse heart, amino acid analysis, 72-73

- molecular shape, 132
- salting-out constant, 274
- sperm whale, molecular dimensions, 132
- Myosin,
 - binding of calcium ions, 599
 - binding of sodium & potassium ions, 521
 - helical configurations, 130
 - lifetime, 49
 - rabbit, amino acid analysis, 71ff.
 - structural relation to keratin, 115
 - sulfhydryl groups, 63

N

- Neon,
 - abundance, 8-10
 - biologically inert, 5
 - comparison with other isoelectronic substances, 28ff.
 - distribution in nature, 4, 8
 - Nickel,
 - abundance, 8-10
 - association constants with bases, 634
 - biological significance, 6
 - complexes with bases, 626, 639, 641
 - distribution in nature, 4, 8
 - percentage in sea water, 21
 - Ninhydrin, 69
 - amino acid reaction, 66ff.
 - Nitrate, standard free energy of formation, 237
 - Nitrite, standard free energy of formation, 237
 - p*-Nitroaniline, dipole moment, 338
 - Nitrogen,
 - absorption coefficient in water, 556
 - abundance, 8-10
 - biological significance, 5
 - distribution in nature, 4, 8, 10-17, 21
 - percentage in sea water, 21
 - Nitromethane, dipole moment, 338
 - Nonaqueous systems, acid-base equilibria, 408
 - Nonelectrolytes, salting-out constants, table, 274
 - Nonionic compounds, solubilities, 45
 - Nonpolar molecules, *see under* Molecules.
 - Nonpolar side chains, protein structure, 89
 - Nucleic acids, 47
 - binding to alkaline earths, 591ff.
 - Nucleoproteins, 54
 - formation from proteins & nucleic acid, 593
 - Nucleosides, 467, 470
 - apparent dissociation constants, 467
 - Nucleotides, 467, 470
 - apparent dissociation constants, 467
- O
- Ocean,
 - composition, 17ff., 23
 - concentration, variations, 23
 - evolution, 17ff., 22
 - temperature regulation, 22
 - Onsager theory of dielectrics, 347ff., 354-356, 362
 - conductivity, 401
 - Optical polarizability, 337
 - Optical rotation, peptides & proteins, 129-131
 - Orientation, molecules in electric field, 332ff.
 - Ornithine, in Gramicidin S, 78
 - Ovalbumin, egg white,
 - acid-base equilibrium, 530
 - amino acid analysis, 72-73
 - C-terminal residues, 84
 - carbohydrate residue as part of molecule, 52
 - dielectric properties, 372
 - helical configurations, 130
 - phenolic hydroxyl groups, 62, 88, 530
 - salting-out constant, 274
 - sulfhydryl groups, 88
 - titration curve, 530
 - Oxalacetate, standard free energy of formation, 237
 - Oxalate ion, complexes with metal ions, 639
 - Oxonium ion, *see* Hydrogen ion.
 - Oxygen,
 - absorption coefficient in water, 556
 - abundance in nature, 8
 - biological significance, 2, 5, 12-17
 - combination with hemoglobin, 607, 635, 653ff.
 - dipole moment, 338

distribution in nature, 3-4, 7-12, 17-18
formation in primitive atmosphere, 14ff.
isotope, O^{18} , for study of hydration of CO_2 , 583
salting-out constant, 274

P

Partial molal heat capacities, *see* Heat capacities.
Partial molal quantities, *see* Molal quantities.
Partial molal volumes, *see* Molal volumes.
Partial pressures, *see* Pressures, partial.
Pepsin,
 N-terminal residue, 84
 phosphate cross-links, 88
 specificity, 74
Peptide linkage, 466, 470
 effect on ionization constants, 466
Peptides,
 activity coefficients, 197, 315ff.
 amino acids & peptides as dipolar ions, 54
 bond,
 coplanar configuration, 101
 synthesis,
 free energy changes, 207-210
 standard free energy, 238
 chains, 74ff., 80, 110ff.
 arrangement in protein molecule, 81
 cross-links & loops, 125ff.
 C-terminal residues, 76, 84
 helical configuration, 110ff.
 hydrogen bonds, 101, 106ff.
 interatomic distances & bond angles, 101-104
 N-terminal residues, 76, 82, 84
 pleated sheet structure, 105ff.
 prolyl residues, 104
 resonance, 101
 sequence of amino acid residues, 89ff.
 silk fibroin, 104-106
 spatial configuration, 99ff.
 chelate complexes, 643ff.
 cyclic, 78
 deuterium exchange in peptides & proteins, 127-128

 formation, theoretical, in earth's history, 14
 melting points, 55
 optical rotation, 129-131
 polypeptides, 540ff.
 solubility, 315ff.
 titration curves, 540ff.
Performic acid, oxidation,
 of disulfide links, 85
 of tryptophan and methionine, 85
pH,
 calculation in systems of known acid strength, 414ff.
 concept, *see* Hydrogen ionactivity, pH
 determination from cells with liquid junctions, 437-450
 ionic strength, effect of, for acid-base pairs of different charge types, 441ff.
 measurements,
 calibration, 444ff.
 glass electrode, 449ff.
 standard solutions, 446-449
 table of values, 448
 standard solutions, 448
Phase, definition, 137
Phase equilibria,
 degrees of freedom, 176-179
 and one-component systems, 177-180
Phase rule, 176
Phenol(s),
 dipole moment, 338
 ionization constants, 426-429, 464
Phenolic hydroxyl groups (*see also* Tyrosine),
 heat and entropy of ionization, 464, 533
 ionization constants, 463-464, 531, 533, 536, 542-545
 role, in protein structure, 62, 88-89, 531, 533
Phenylalanine,
 constituent in important proteins, 72
 structure, 58
D-Phenylalanine in Gramicidin S, 78
Phenylisothiocyanate, reagent for determination of N-terminal amino acids, 83
Phenylthiohydantoin, formation from N-terminal amino acids, 83
Phenylthiourea, salting-out constant, 274

- Phosphate(s),
 cross-links in proteins, 86-88
 ionization constants, 407, 442-443, 452, 467, 470
- Phosphoglyceric acid, in photosynthesis, 551
- Phosphoric acids,
 acid dissociation constants, 407, 452, 456
 derivatives, dissociation constants, 452, 467, 470
- Phosphorus,
 biological significance, 3, 5ff.
 distribution in nature, 3-4
 percentage in sea water, 21
- Photosynthesis,
 introduction of oxygen into primitive atmosphere, 15
 phosphoglyceric acid, 551
 ribulose diphosphate, 551
 role of carbon dioxide, 551
- Pituitary hormones,
 corticotropins, structure, 95
 melanotropins, structure, 95
 oxytocin, structure, 78-79
 vasopressins, 79-80
- pK_A values, *see* Acid dissociation constants.
- Planck's formulation of Second Law of Thermodynamics, 147
- Plasma, *see under* Blood.
- Pleated sheet structure, 105ff.
 peptide chains, antiparallel, 105-106
 parallel, 108ff.
- Point dipole, definition, 257ff.
- Poisson's equation, 249ff.
 Debye-Hückel theory, relationship, 285ff.
 spherical symmetry, 250ff.
- Polar coordinates, relation to Cartesian coordinates, 245
- Polar groups, *see* Dipole moments.
- Polar liquids (*see also* Dipole moments),
 dielectric constants, theory, 341ff.
 dipolar ions in solution, 367ff.
 experimental studies, comparison with theory, 364ff.
 water, dielectric constant, 43, 367
- Polarizability, electrical,
 electronic, 337
 hydroxydecanoic acid polymers, 340
 molar, 334
 Kirkwood's theory, 362
 molecular,
 definition, 334
 Kirkwood's theory, 362
 Onsager theory, 348
 optical, 337, 346
 per unit volume, 334ff.
 in polar liquids, relation to dielectric constant, 364ff.
 relation to molecular properties, 328ff.
 total, relation to dielectric constants in Kirkwood's theory, 361
- Polarization (*see also* Polarizability), 328ff.
- Poly-L-alanine, helical structure, 129
- Polyamino acids, *see* Peptides, polypeptides.
- Polybasic acids, 481ff., 488ff.
 electrostatic effects on ionization, 512ff.
 microscopic constants, 489ff.
 titration constants, 481ff.
- Polyelectrolytes, 408
 fibers as means for transforming chemical energy into mechanical work, 542
 molecular configurations, possibilities, 541
 titration curves, 540ff.
- Polyglycine,
 hydrogen bonds, 120-121
 polyglycine II, molecular structure, 119-121
- Polypeptides, *see* Peptides.
- Poly-L-proline, molecular structure, 122
- Polystyrene sulfonate, ion exchange resin, 68
- Poly-L-tyrosine, 542ff.
 ionization, effect of ionic strength, 542-543
 titration curve, 543
- Polyvalent acids (*see also* Acids; Acid dissociation constants),
 association constants, 518ff.
 charged sphere model, activity coefficients, 514ff.
 dielectric sphere model with fixed basic sites, 546

- equilibrium between different ionic forms, 510ff.
general relations, 487-494
hydrogen bonding, effects on titration, 545
Linderstrøm-Lang equation for titration of polyvalent acids, 515ff.
 n equivalent & independent acid groups, 493
standard deviation of net charge, 511
titration curves, methods of calculating, 515ff.
- Polyvalent ampholytes, 506-510
 general equations, 506ff.
 isoelectric points, 507ff.
- Potassium,
 biological significance, 5ff.
 distribution in nature, 3-4, 9-10
 percentage in sea water, 21
- Potential, liquid junction, 439ff.
- Potential energy of a dipole, 254, 297ff.
- Porphyryns,
 chelate complexes, 609
 role in earth's early history, 17
- Potentials, chemical, *see* Chemical potentials,
 gravitational, 245-246
- Potentials, electrical, 243ff.
 charged spheres, 252ff.
 dipoles, 254ff., 297ff.
 gradient, 244
 electric intensity, 245
 interionic forces, 282-296
 ionic solutions, 282-296
 Poisson's equation, 249-253, 286ff.
 self-potential of sphere, definition, 253
 sphere in dielectric medium, 344
- Potentiometer, emf measurements, 430ff.
- Poynting's law, energy flux, 323
- Pressure, partial, definition, 182-183
- Pressure volume work, 162ff.
- Proline (*see also* Polypyrroline),
 acid dissociation constants, 452
 activity coefficient, 197
 constituent in important proteins, 72
 hydroxyproline & proline in collagen structure, 122-123
 peptide chains, 104, 122, 131
 residues, influence on helical structure stability, 125
 role in collagen, 116-117
 structure, 59, 68ff.
- n -Propanol, dipole moment, 338
 standard free energy of formation, 237
 thermodynamic properties, aqueous solutions, 187-194
- Propionic acid, acid dissociation constant, 452
- n -Propyl alcohol, *see* n -Propanol.
- Proteins (*see also* Peptides, and individual proteins),
 acid-base equilibria, 522-540
 amino acids,
 analysis, 65, 72-73
 structure, 58-59
 binding of ions & molecules to proteins, 521, 591ff.,
 biosynthesis, 49
 carbamino derivatives, 575
 complexes with lipides, 52
 cross-links & loops, 125ff.
 crystals, 100, 131, 320
 C-terminal residues, 76, 84
 cysteine sulfhydryl groups, 63
 denaturation, 51, 64, 84-89, 531, 537-540
 deuterium exchange in peptides & proteins, 127-128
 disulfide cross-links, 48, 75ff., 85ff., 96-97, 125ff.
 enzymes, 48ff.
 extraction from living tissues, 50
 globular,
 helical & nonhelical regions, 129
 X-ray diffraction patterns, 131-132
 hair proteins, 47ff.
 hydrolysis with acids, alkali, & enzymes, 56, 65
 immunological reactions, 51, 593
 ion binding, calculated from pH shifts, 599-603
 ion mobilities, lower than in small ions, 400
 ions removed from, by electrodialysis & ion exchange resins, 600
 keratin, 48ff.
 molecular complexes, 50, 591ff.
 molecular weights, 72
 minimum, 80
 N-terminal residues, 76, 82, 84

- nonpolar side chains, 64, 89
 optical rotation, 129-131
 phosphate cross-links, 86ff.
 reaction with CO_2 to form carbamates, 571-578
 purity, 51
 salting out, 275ff.
 salting-out constants, table, 274
 solubility as criterion of purity, 51
 stability, in living tissues, 50
 structure, 47ff., 94, 98-132
 sulfhydryl groups, 63
 tendon proteins, *see* Collagen
 wool proteins, *see* Keratin
 Proton acceptors, *see* Bases.
 Proton donors, *see* Acids.
 Proton transfer,
 in acid-base equilibria, 406ff., 477ff.
 in liquid water, 34ff.
 Proton unit of charge, definition, 28, 243
 Purines, 467, 470
 apparent dissociation constants, 467
 Purity of proteins, criteria, 51
 Pyridine,
 complexes with metal ions, 640
 dipole moment, 338
 nucleotides, complexes with enzymes, 592, 619-620
 Pyrimidines, 467, 470
 apparent dissociation constants, 467
 Pyrophosphate, bridges in proteins, 87
 Pyruvate ion, standard free energy of formation, 237
 Pyruvic acid,
 ionization constant, 466
 carbon dioxide metabolism, 551
- Q**
- Quasi-static processes 145-146
- R**
- Raman spectra, 55
 acids & conjugate bases, 425
 Randomness, entropy, distribution of molecules, 225ff.
 Raoult's law, 184ff.
 negative deviations, 188
 positive deviations, 187
 Reaction field, *see* Onsager theory; Dipole moments.
 Reactions, rapid,
 measurement, 581-583
 rate and heat of reaction for dehydration of H_2CO_3 , 581-583
 rate of hydration of CO_2 , 583-587
 Reciprocal plot of association data, 616-617
 Relaxation time, electrical conductivity, 401-402
 Residual entropy, *see under* Entropy.
 Resistance, electrical, 386-387
 specific resistance, 386-387
 Resonance,
 carbonic acid & carbonate ion, 554
 effect on acid strengths, 463
 frequency, condenser circuit, 382
 Reversible processes, 148ff.
 maximum electrical work in, 431
 Ribonuclease, beef pancreas,
 acid-base equilibria, 536-537
 amino acid analysis, 70-73
 amino acid sequences, 92, 98-99
 disulfide cross-links, 86, 96-99, 125-126
 helical and nonhelical configurations, 130
 isoelectric point, 537
 molecular structure, 97-99, 128-129
 performic acid oxidation, 86
 phenolic hydroxyl groups in structure, 88, 129, 537
 purification, 52
 stereochemistry, 128-129
 table, comparison of titration data & amino acid analysis, 536
 Ribulose diphosphate, photosynthesis, 551
 Rocks, weathering by action of water, 19-20
 Rubey, W. W., theories on primitive atmosphere, 15
 Rubidium, percentage in sea water, 21
- S**
- Salicylic acid, ionization constant, 463-464, 467, 531

- Salting-in constants,
 amino acids, 316
 dipole moments, relation, 312ff.
 table, 316
 peptides, 316
- Salting-in effect,
 dipolar ions, 264
 dipole moments, relation, 314-315
 hydrocyanic acid, 264
 hydrogen cyanide, 271
 solutions, 310ff.
 theory, 296-310
- Salting-out effect,
 acetone, 271
 chemical potential, relationship, 266ff.
 dipolar ionization, 304-305
 ethanol solutions, 263ff.
 ether solutions, 263ff.
 protein solutions, 264, 275ff.
 salting-out constants, table, 274
 theory, 263-282
 vapor pressure, relationship, 270ff.
- Sanger's procedure, peptide chain sequences, 89-93
- Sanger's reagent, *see* Dinitrofluorobenzene.
- Sarcosine, activity coefficient, 197
- Scalar products, of two vectors, 359ff.
- Scandium,
 biological significance, 5
 distribution in nature, 4
 percentage in sea water, 21
- Scatchard plot of association data, 616-617
- Sea water (*see also* Ocean),
 elements in composition, 20ff.
 ionic composition, 17-20
 pH, 23
- Sedimentation equilibrium, relation to molecular weight, 219-224
- Selenium, percentage in sea water, 21
- Self-potential of sphere, *see under* Potentials, electrical.
- Semiquinones in oxidation reactions, 606
- Serine,
 acid dissociation constants, 452
 activity coefficient, 197
 constituent in important proteins, 72
 in phosphate linkages, 86-88
 structure, 58
- Serum albumin,
 binding of anions, 521
 cadmium ions, interaction, 643
 cattle,
 acid-base equilibria, 532-536
 amino acid analysis, 72-73, 76
 isoelectric point, 599
 N-terminal residue, 84
 complexes with anions, 594-603, 645-653
 disulfide cross-links, 86
 disulfide dimer, 63
 dodecylsulfate, methyl orange,
 salicylate, 652
 helical configurations, 130
 human, 76
 interaction with chloride & thiocyanate ions, 597, 646ff.
 with anions, pH changes associated, 600-601
 with molecules & ions, 591ff., 608
 intramolecular alteration, in acid solutions, 535-536
 lead (Pb II) ions, interaction, 643
 mercury dimer, 63
 mercury salts, 63-64
 molecular dimensions, 512
 phenolic hydroxyl groups, 89
 entropy of ionization, 533
 heat of ionization, 533
 stabilization against denaturation, 89
 titration curve, 532
 effect of ionic strength, 533
 zinc-ion interaction, 642
- Serum mercaptalbumin,
 bovine, binding of anions, 650
 dielectric properties, 372
 interaction with mercury 63, 645
- Silicon,
 abundance, 8-10
 biological significance, 5
 comparison of properties with those of carbon, 24
 distribution in nature, 4, 8
 percentage in sea water, 21
- Silk fibroin, *see under* Fibroin.
- Silver,
 association constants with organic bases, 634, 639

- percentage in sea water, 21
- Sodium,
 - biological significance 3, 5ff.
 - distribution in nature, 3-4, 9
 - percentage in sea water, 21
- Solid angle, definition, 248
- Solubilities, Solubility,
 - activity coefficients, determining, by
 - solubility measurements, 200, 308-320
 - criterion of protein purity, 51
 - dipolar ions, 55, 306ff.
 - distribution coefficients & solubilities in relation to activities, 198ff.
 - increased, with decreased activity coefficients, 282ff.
 - uncharged isomers of dipolar ions, 306
 - various substances, 45
- Solutions,
 - ideal or perfect, 184-185, 229-231
 - non-ideal, activity coefficients, 187-200
(*See also Chapters 5, 8, 9, and 11, and references under activity coefficients.*)
- Somatotropin, growth hormone,
 - C-terminal and N-terminal residues, 84
- Space charge (*see also* Poisson's equation), 249ff.
 - calculation, in ionic solution, 293ff.
- Specific conductance, 386-387
- Specific heat (*see also* Heat capacity),
 - water and other substances, 37ff.
- Specific resistance, 386-387
- Spectrophotometry,
 - determination of ionization constants, 423-429, 531, 533, 537, 543-545
 - study of dye binding to proteins, 596
- Spectroscopic measurement, study of molecular complex formation, 596
- Sphere, charged, *see* Charged sphere.
- Standard entropy of chemical substances, 234
 - of formation
 - of ionization of acids, 451-457, 463-464, 533
- Standard free energies of formation, 234
 - table of data, 236-237
- Standard potentials,
 - galvanic cells, 433ff.
 - half-cells, 434ff.
- Standard solutions for calibrating pH measurements, 446-449
 - table of values, 448
- Standard states,
 - activity coefficients, 189-194
 - conventions for choosing, 192
 - definition, state of unit activity, 182
 - relation to activity, 182
- Starch, 48
- Statistical factors,
 - acid ionization, 485, 487-494
 - association reactions, 624, 630, 635-636, 656-657
- Statistical interpretation of entropy, 224-232
- Stokes' law, resistance to motion of a sphere in a viscous liquid, 391, 402
- Strontium,
 - biological significance, 6
 - percentage in sea water, 21
- Subtilisin,
 - proteolytic enzyme from *Bacillus subtilis*, 93
 - use in study of protein structure, 93
- Succinate ion, standard free energy of formation, 237
- Succinic acid,
 - acid dissociation constants, 452, 456
 - standard free energy of formation, 237
- Sucrose, standard free energy of formation, 237
- Sulfanilamide, carbonic anhydrase inhibitor, 588
- Sulfhydryl compounds, interactions with metal ions, 644
- Sulfhydryl groups (*see also* Cysteine),
 - amino acids, 59
 - ionization 465, 468, 497ff.
 - proteins, 48, 62-64
- Sulfur (*see also* Cysteine, Cystine, Methionine),
 - abundance, 8-10
 - biological significance, 5
 - distribution in nature, 3-4, 8
 - occurrence in sea water, 20
 - percentage, 21
- Sulfur-containing R groups, amino acid structure, 59
- Sulfur dioxide, absorption coefficient in water, 556

- Surface, equipotential, *see* Equipotential surfaces.
- Surface charges, electric field, 324ff.
- Surface tension,
 heat of vaporization, relation, 42ff.
 table, 42
- Symmetrical dibasic acids,
 equations, 482
 with independent groups, 483-484
- System, Systems,
 centrifugal fields, 219
 closed, 137ff.
 components, definition, 180
 gravitational fields, 219
 homogeneous, 137ff.
 isolated, 137ff.
 macroscopic, corresponding microscopic configurations, 225-226
 thermodynamic, 137ff.
- T**
- Temperature, 138ff.
 absolute, 148, 153ff.
 concepts, 151
 dielectric constant, dependence on, 328
 relation to empirical temperature, 166-167
 relation to irreversible processes, 157ff.
 dimensions, 141-142
 empirical, 154
- Tension, surface, *see* Surface tension.
- Thermodynamics (*see also* Activities; Energy; Temperature; Chemical Potentials), 136ff.
 concepts, application to irreversible processes, 217
 energy levels of molecules, 227
 First Law, 142-146
 Functions, equations relating, 163ff.
 Second Law, 146-157
 thermodynamic system, microscopic configurations, 225-226
 Third Law, 155, 164, 228ff.
- Thiocyanate ions,
 association with mercaptalbumin, 651
 binding to serum albumin, 597ff., 646ff.
- Thioglycolic acid, ionization constants, 465, 498ff.
- Thiol compounds, *see* Sulfhydryl compounds.
- Threonine,
 acid dissociation constants, 452
 activity coefficient, 197
 constituent in important proteins, 72
 L-threonine, standard free energy of formation, 237
 phosphate cross-links, 87
 structure, 58, 61
- Thymine, apparent dissociation constant, 467
- Thyroxine, structure, 58
- Tin, percentage in sea water, 21
- Titanium,
 biological significance, 5
 distribution in nature, 4
- Titration constants,
 in association reactions, definition, 625
 individual group of *n*-valent acid, 488
 polybasic acids, 481ff.
 relations to dissociation constants, 482, 492-494
- Titration curves (*see also* Acid dissociation constants), 419, 477ff., 513-547
 effect of ionic strength, 513-547
 hemoglobin, 525, 537
 β -lactoglobulin, 523
 other proteins, 540
 ovalbumin, 530
 polyelectrolytes, 542
 ribonuclease, 536
 serum albumin, 533
- Toluene, dipole moment, 338
- Transference numbers, 393-395
 Hittorf method of determining, 394-395
 moving boundary method of determining, 395-401
- Trichloroacetate, association with mercaptalbumin, 651
- Triglycine (Glycylglycylglycine),
 activity coefficient, 197
 solubility,
 water & ethanol, 306
- Trimethylamine, dipole moment, 338
- Trimethylammonium ion, acid dissociation constants, 452

- 2,2,3-Trimethylbutane, dipole moment, 338
- N-Trimethylcysteine, *see* Cysteine betaine.
- Tris(hydroxymethyl) aminomethane, acid dissociation constants, 452
- Tropomyosin, helical configurations, 130-131
- Trypsin,
cattle, C-terminal & N-terminal residues, 84
inhibitor, 84
pancreatic, 84
soybean, 84
specificity, 75
protein structure study, 92
- Trypsinogen, cattle, N-terminal residue, 84
- Tryptophan,
constituent in important proteins, 72
oxidation by performic acid, 85
structure, 58
- Tyrosine,
apparent dissociation constants, 464
constituent in important proteins, 72
copolymers with alanine, aspartic acid, lysine, 544-545
ionic strength, effect on titration curves, 544-545
hydroxyl groups, 88, 128
ionization constants, 530-531
L-tyrosine, standard free energy of formation, 237
microscopic ionization constants, 504
O-methyl tyrosine, microscopic ionization constants, 504
ovalbumin structure, role of tyrosine, 62, 531
ribonuclease structure, role of tyrosine, 88-89, 537
salting-out constant, 274
serum albumin, role of tyrosine, 533
structure, 58
- U
- Ultracentrifuge,
molecular complex studies, 598, 619-620
molecular weight determination, sedimentation equilibrium, 223
- Ultrafiltration in study of molecular complexes, 595
- Ultraviolet absorption spectra,
cysteine, 499-503
glycyl-L-tyrosine, 426-428
ovalbumin, 530-531
ribonuclease, 537
sulfhydryl compounds, 499-503
tyrosine derivatives, 426ff; 531, 533, 537
- Unshared electron pairs,
presence characteristic of bases, 408
relation to water and ice structure, 30ff., 35-36
- Uracil, apparent dissociation constant, 467
- Urea,
action on insulin and ribonuclease, 128
denaturing action on proteins, 64, 124
dipole moment, 338
disrupting action on helical structures, 124
solvent for glycine and peptides, 371, 372
standard free energy of formation, 237
- Urease, 64
- Urey, H. C., theories on primitive atmosphere, 15
- Uridine, apparent dissociation constant, 467
- V
- Valeric acid, apparent dissociation constant, 466
- Valine,
activity coefficient, 197
constituent in important proteins, 72
structure, 58
- Vanadium,
biological significance, 5
distribution in nature, 4
percentage in sea water, 21
- Van't Hoff's equation, heat of reaction and equilibrium constant, 203
- Van Vleck "4 π I/3 catastrophe,"
dielectric constants, general theory, 342, 364
- Vapor pressure, 182-183
activity coefficients, 191-194
relation to thermodynamic activity, 181

salting-out effect, relationship, 270ff.
Vaporization, heat of, *see* Heat of vaporization.
Vector(s),
 additivity of polar groups, 339
 product,
 definition, 255-256
 dipole moments, 352
 quantities, notation, 243
 scalar products, 359ff.
Vibrational frequencies, dipolar ions, 55
Viscosity, ion mobilities, relationship, 390-391
Vitamin A, visual purple in retina, 54
Vitamin B-12, structure, 609-610
Volume, changes on ionization, 456
 partial molal, 170-174

W

Wheatstone bridge, measurement of dielectric constants, 384
Wien effect,
 intense electric fields, 404
 ionization of carbonic acid, 559
Wool, 47-48
Work,
 change of length of elastic body, 217-219
 in gravitational or centrifugal field, 219ff.
 maximum electrical work, in galvanic cell, 431ff.
 maximum, relation to free energy, 159-162
 relation to energy, 143-146, 149ff.
 "pressure-volume" work, 160
 "useful" work and free energy, 160
Water (*see also* Ice; Ocean; Sea water),
 as acid & base, 412-413
 acid dissociation constant, 407, 452
 biological significance, 27-46
 comparison with other isoelectronic substances, 27ff.
 covalent bonds, strength, 32ff.
 dielectric constant, 43, 367
 dipole moment, 338
 expansion on freezing 40-41, 166
 heat capacity of water & other substances, 37

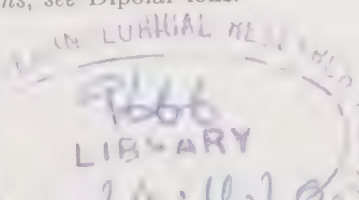
heat of vaporization, 38
latent heat of fusion, 39-40
liquid,
 quasi-crystalline structure, 33, 41
 standard free energy of formation, 237
mixtures with *n*-propanol, activity & vapor pressures, 187-194
molecule,
 electrical polarity, 30ff.
 electronic structure, 30ff.
 triangular shape, 30ff.
phase diagram, 178
proton transfer in liquid, 34ff.
residual entropy at 0° K, 231
solvent for dipolar ions, 55, 306-320
specific heat, 38
surface tension, 42-43
thermal properties, 37ff.
uniqueness of structure, 36
vaporization & change of entropy, 167ff.

X

X-ray diffraction studies,
 collagen, 116ff.
 crystalline proteins, 131
 ice, 31-32
 α -keratin, 114
 β -keratin, 107
 liquid water, 33
 poly-L-alanine, 129
 significance for protein & other structures, 100-101, 131

Z

Zinc,
 association constants with bases, 626-634, 639, 641
 binding to proteins, 592
 biological significance, 6
 chelate complexes,
 tetrahedral, 633
 constituent of carbonic anhydrase enzyme, 6, 588
 distribution in nature, 3-4
 enzyme constituent, 620
 percentage in sea water, 21
Zwitterions, *see* Dipolar ions.



Date 29/4/98
 14.7.80
 m. 8.89
 C. F. T. R. I. LIBRARY, M
 REF
 Acc No. 8666

REF

Please return this publication on or before the last DUE DATE stamped below to avoid incurring overdue charges.

To be issued from 9th January 1971.

CHECKED
2008

VERIFIED
2013



9666
Biophysical chem.



